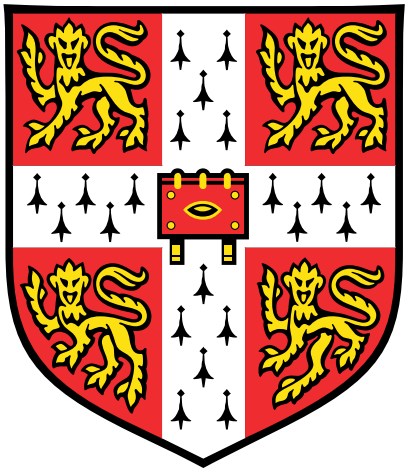
# Treslin and its role in the assembly of the replicative DNA helicase



**Vincentius Aji Jatikusumo**

Department of Biochemistry
University of Cambridge

This dissertation is submitted for the degree of
*Doctor of Philosophy*

Peterhouse August 2019

# Declaration

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgements and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Acknowledgements and specified in the text.

It does not exceed the prescribed word limit for the Degree Committee for the Faculty of Biology. This dissertation contains fewer than 60,000 words — excluding appendices, bibliography, footnotes, tables, and equations — and has fewer than 150 figures.

Vincentius Aji Jatikusumo
August 2019

**Treslin and its role in the assembly of the replicative DNA helicase**

Vincentius Aji Jatikusumo

DNA replication is an essential biochemical process that underlies genetic inheritance. In eukaryotic cells, this process is carried out by a multi-subunit protein complex called replisome. The core of replisome is made up of six-subunit protein ring Mcm2–7, which contains weak DNA helicase activity. In order to fully activate the helicase activity, protein co-activators of Cdc45 and GINS have to be recruited to assemble a replicative DNA helicase called the CMG (Cdc45-Mcm2–7-GINS) complex. The active CMG subsequently unwinds the parental DNA duplex to provide single-stranded DNA templates for replicative DNA polymerases. The precise mechanism for CMG assembly and activation is still poorly understood. Recent studies using model organisms suggested the importance of protein called Sld3/Treslin for CMG assembly.

The aim of the project was to understand in detail how human Treslin participates in the assembly of the CMG helicase. The interactions of Treslin with Cdc45, Mcm2–7, and DNA were characterised by a series of biochemical, biophysical, and structural biology experiments.

Cdc45–Treslin interaction was shown to be conserved from budding yeast to human, in which it is mediated by a domain called Sld3/Treslin Homology Domain (SHD). However, human Treslin contains an additional Cdc45-binding site — next to the C-terminus of SHD — in an intrinsically disordered region referred to as Treslin 'extension' ($Treslin_{ext}$). $Treslin_{ext}$ was also demonstrated to interact with Mcm2–7 ring and DNA. S-phase Cyclin-Dependent Kinase (S-CDK) appeared to regulate Treslin's interaction with Cdc45 and Mcm2–7, in which it targets a number of potential S-CDK phosphorylation sites present in $Treslin_{ext}$. As a part of this project, low-resolution Cdc45–Treslin and medium-resolution of MCM single hexamer 3D maps were obtained through single-particle cryo-electron microscopy (cryo-EM) analyses.

The results provide molecular basis for Treslin's role in the assembly of the CMG complex. Treslin acts as a molecular chaperone for Cdc45 recruitment to the DNA-bound Mcm2-7 ring. Treslin's role during CMG assembly may involve its binding ability to the Mcm2–7 and DNA. Furthermore, the regulatory role of S-CDK in Treslin's interactions with Cdc45 and Mcm2–7 suggests a link of the CMG assembly control to the cell cycle in human.

# Acknowledgements

I would like to express my gratitude to my supervisor Professor Luca Pellegrini, for giving me the opportunity to join his laboratory and his support throughout the project.

I am thankful to Dr Neil Rzechorzek and Dr Joseph Maman for all their exceptional advice and assistance. I am grateful for helpful discussions from the present or past members of Pellegrini lab: Dr Mairi Kilkenny, Dr Sandro Holzer, Dr Daniel Bollschweiler, Dr Laura Radu, Dr Luay Joudeh, Dr Johanna Syrianen, Dr Aline Simon, Dr Yue Zhang, and Dr Christopher Morton.

Advice given by my first year examiners, Professor Sir Tom Blundell and Dr Philip Zegerman, has been extremely helpful for the completion of this PhD.

I wish to acknowledge the technical assistance given by Dr Dimitri Chirgadze, Dr Steven Hardwick, and Lee Cooper at the Cryo-EM facility at the Department of Biochemistry, Cambridge. Thanks to Dr Mark Skehel and Dr Sarah Maslen at Laboratory of Molecular Biology, Cambridge for the HDX-MS experiments. Lastly, I want to acknowledge all the staff at the DNA sequencing facility at the Department of Biochemistry, Cambridge and eBIC Diamond Light Source, Oxfordshire.

I would like to express my great appreciation to the Gates Cambridge Scholarship and its inspiring group of scholars. Their help and inspiration enabled me to complete this journey.

My thankfulness to my sister Xaveria Nitra Larasati, grandma Eyang Sri Koesmiati (who believed in me before everyone else), and the family of Pakde Dr Bambang Eko Afiatno.

My gratitude to my partner, Dr Nikola Doležalová, who keeps me going in the toughest of times with her unending inspiration, patience, and support. This will not be possible without you.

Lastly, but definitely not the least, my deepest gratitude and eternal love to my parents, dr Dwi Rini Indradjajati and dr Eric Soeroso. Their endless love, support, understanding, and perseverance inspire me all the time. Thank you for showing and letting me to be independent. I dedicate this work to them.

# Abstract

DNA replication is an essential biochemical process that underlies genetic inheritance. In eukaryotic cells, this process is carried out by a multi-subunit protein complex called replisome. The core of replisome is made up of six-subunit protein ring Mcm2–7, which contains weak DNA helicase activity. In order to fully activate the helicase activity, protein co-activators of Cdc45 and GINS have to be recruited to assemble a replicative DNA helicase called the CMG (Cdc45-Mcm2–7-GINS) complex. The active CMG subsequently unwinds the parental DNA duplex to provide single-stranded DNA templates for replicative DNA polymerases. The precise mechanism for CMG assembly and activation is still poorly understood. Recent studies using model organisms suggested the importance of protein called Sld3/Treslin for CMG assembly.

The aim of the project was to understand in detail how human Treslin participates in the assembly of the CMG helicase. The interactions of Treslin with Cdc45, Mcm2–7, and DNA were characterised by a series of biochemical, biophysical, and structural biology experiments.

Cdc45–Treslin interaction was shown to be conserved from budding yeast to human, in which it is mediated by a domain called Sld3/Treslin Homology Domain (SHD). However, human Treslin contains an additional Cdc45-binding site — next to the C-terminus of SHD — in an intrinsically disordered region referred to as Treslin 'extension' ($\text{Treslin}_{\text{ext}}$). $\text{Treslin}_{\text{ext}}$ was also demonstrated to interact with Mcm2–7 ring and DNA. S-phase Cyclin-Dependent Kinase (S-CDK) appeared to regulate Treslin's interaction with Cdc45 and Mcm2–7, in which it targets a number of potential S-CDK phosphorylation sites present in $\text{Treslin}_{\text{ext}}$. As a part of this project, low-resolution Cdc45–Treslin and medium-resolution of MCM single hexamer 3D maps were obtained through single-particle cryo-electron microscopy (cryo-EM) analyses.

The results provide molecular basis for Treslin's role in the assembly of the CMG complex. Treslin acts as a molecular chaperone for Cdc45 recruitment to the DNA-bound Mcm2-7 ring. Treslin's role during CMG assembly may involve its binding ability to the Mcm2–7 and DNA. Furthermore, the regulatory role of S-CDK in Treslin's interactions with Cdc45 and Mcm2–7 suggests a link of the CMG assembly control to the cell cycle in human.

# Table of contents

# List of figures

# List of tables

# List of Abbreviations

**Acronyms / Abbreviations**

| | |
|---|---|
| 2×YT | Yeast Extract Tryptone |
| 6A | 6 Alanine mutants |
| Å | Ångström unit = 0.1 nm |
| AAA+ | ATPases Associated with diverse cellular Activities |
| ACS | ARS-Consensus Sequence |
| And1 | Acidic Nucleoplasmic DNA-Binding Protein 1 |
| APC/C | Anaphase Promoting Complex/Cyclosome |
| ARS | Autonomous Replication Sequence |
| ATP | Adenosine Triphosphate |
| BirA | Bifunctional Ligase/Repressor A |
| BLI | Bio-Layer Interferometry |
| bp | base pair |
| BRCT | Breast Cancer Protein 1(BRCA1) C-terminal |
| BSA | Bovine Serum Albumin |
| Cdc45 | Cell Division Cycle 45 protein |
| Cdc6 | Cell Division Cycle 6 protein |
| CDK | Cyclin-Dependent Kinase |

| | |
|---|---|
| cDNA | complementary DNA |
| Cdt1 | Chromatin Licensing and DNA Replication Factor 1 |
| Chk1 | Checkpoint Kinase 1 |
| CID | CMG-Interacting Domain |
| CIT | Conserved In Treslin |
| CMG | Cdc45, Mcm2–7, GINS |
| CMV | Cytomegalovirus |
| Cryo-EM | cryo electron microscopy |
| CTF | Contrast Transfer Function |
| Ctf4 | Chromosome Transmission Fidelity Protein 4 |
| CV | Column Volume |
| DDK | Dbf4-Dependent Cdc7 Kinase |
| DHH | Desert Hedgehog phosphoesterase |
| DNA | Deoxyribonucleic Acid |
| Dpb11 | DNA Polymerase B subunit 11 |
| DPC | DNA-Protein Crosslink |
| dRI | differential Refractometer Index |
| DSB | Double Strand DNA Break |
| dsDNA | double-stranded DNA |
| DTT | dithiothreitol |
| eBIC | Electron Bio-Imaging Centre |
| EDTA | ethylenediaminetetraacetic acid |
| EMDB | Electron Microscopy Data Bank |
| EMSA | Electrophoretic Mobility Shift Assay |

| | |
|---|---|
| ext | extension region on Treslin (Treslin$_{ext}$) |
| FACT | Facilitates Chromatin Transcription |
| FAM | Fluorescin Amidite |
| FBS | Fetal Bovine Serum |
| G4 | G-Quadruplex |
| GINS | Go-Ichi-Ni-San, Japanese for 5-1-2-3 (Sld5-Psf1-Psf2-Psf3) |
| GST | Glutathione S-transferase |
| HDX-MS | Hydrogen Deuterium Exchange - Mass Spectrometry |
| HEK-293F | FreeStyle$^{TM}$ Human Embryonic Kidney 293 mammalian cells |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HR | Homologous Recombination |
| HRP | Horseradish Peroxidase |
| I.M.A.G.E | Integrated Molecular Analysis of Genomes and their Expression |
| ICL | Interstrand Crosslink |
| IGEPAL | octylphenoxylpolyethoxyethanol |
| IPTG | Isopropyl $\beta$-D-1-thiogalactopyranoside |
| IUPred | Intrinsically Unstructured Protein Prediction |
| $K_d$ | Dissociation constant |
| kV | kilo volt |
| LB | Lysogeny Broth |
| LDS | Lithium Dodecyl Sulphate |
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionisation-Time of Flight mass spectrometry |
| MBD | MTBP-Binding Domain |

| | |
|---|---|
| MBP | Maltose-Binding Protein |
| MCM | Mini-Chromosome Maintenance |
| MCM-DH | MCM double hexamer |
| MCM-SH | MCM single hexamer |
| MCS | Multiple Cloning Site |
| MMR | Mismatch Repair |
| MOI | Multiplicity of Infection |
| Mrc1 | Mediator of the Replication Checkpoint 1 |
| MTBP | Mdm2-Binding Protein |
| Mw | Molecular Weight |
| Ni-NTA | Nickel-Nitrilotriacetic Acid |
| NTD | N-Terminal Domain |
| OCCM | ORC, Cdc6, Cdt1, Mcm2–7 |
| OD | Optical Density |
| ORC | Origin Recognition Complex |
| PBS | Phospho-Buffered Saline |
| PCNA | Proliferating Cell Nuclear Antigen |
| PDB | Protein Data Bank |
| PEI | Polyethylenimine |
| PIP | PCNA-Interacting Protein |
| Pre-IC | Pre-Initiation Complex |
| Pre-RC | Pre-Replication Complex |
| Rad53 | Radiation sensitive protein 53 |
| RELION | Regularised Likelihood Optimisation |

| | |
|---|---|
| RER | Ribonucleotide Excision Repair |
| RNA | Ribonucleic Acid |
| RPA | Replication Protein A |
| RPC | Replisome Progression Complex |
| S-CDK | Synthesis Phase (S-phase) Cyclin-Dependent Kinase |
| SDS-PAGE | Sodium Dodecyl Sulphate - Polyacrylamide Gel Electrophoresis |
| SEC-MALS | Size-Exclusion Chromatography - Multi-Angle Light Scattering |
| Sf9 | *Spodoptera frugiperda* 9 insect cells |
| SHD | Sld3/Treslin Homology Domain |
| Sld2 | Synthetic Lethality with Dpb11, protein 2 |
| Sld3 | Synthetic Lethality with Dpb11, protein 3 |
| SOC | Super Optimal broth with Catabolite repression |
| SPR | Surface Plasmon Resonance |
| ssDNA | single-stranded DNA |
| T/S-P | threonine/serine-proline motif, CDK phosphorylaton site |
| TBE | Tris-Borate EDTA |
| TEMED | Tetramethylethylenediamine |
| TEV | Tobacco Etch Virus |
| TICCR | TopBP1-Interacting Checkpoint and Replication Regulator (also known as Treslin) |
| Tof1 | Topoisomerase Interacting Factor 1 |
| TopBP1 | Topoisomerase II Binding Protein 1 |
| TRCT | Treslin C-Terminal domain |
| Treslin | TopBP1-Interacting, Replication-Stimulating Protein (also known as TICCR) |

| | |
|---|---|
| tRNA | Transfer Ribonucleic Acid |
| VPP | Volta Phase Plate |
| WT | Wild-type |

# Chapter 1

# General Introduction

## 1.1 DNA replication is essential for genetic inheritance

In all living cells, hereditary information is stored in the form of complementary double-stranded DNA (dsDNA). This information contains specific instructions for various cellular activities, from growth to programmed death. The central dogma of molecular biology and genetics succinctly defines the steps needed to decipher these genetic instructions: DNA is first transcribed to messenger RNA (mRNA), and mRNA is then translated into protein (figure 1.1). Proteins — which vary in terms of size, structure, and function — carry out most of the cellular processes that define our individual characteristics.

In order to pass down this genetic information across generations, all organisms need to duplicate faithfully the complete set of DNA instructions (genome) in each parental cell. A copy of these instructions is then transferred to each of two identical offspring cells. This strict requirement is achieved by a biochemical process called DNA replication. Briefly, it is defined as a polymerisation process that produces an exact copy of DNA molecule based on the pre-existing parental template. Considering its fundamental importance, DNA replication can be considered as a universal and essential feature of all organisms.

DNA replication has to be complete, highly accurate, and must happen only once per cell division. To meet these strict criteria, the cell has evolved a complex and highly regulated protein machine called the replisome. Broadly, the replisome consists of a core of proteins involved in DNA unwinding and synthesis supported by a range of processivity, protection, and scaffolding factors (Zhang and O'Donnell, 2016). The replisome's activity is regulated by several checkpoint proteins at various stages of the cell cycle. Disruption of this system has been shown to cause accumulation of mutations across the genome, which can result in several human pathologies such as cancer, neurodegenerative conditions, and early ageing (Barlow and Nussenzweig, 2014; Bicknell et al., 2011; Hills and Diffley, 2014; Jackson

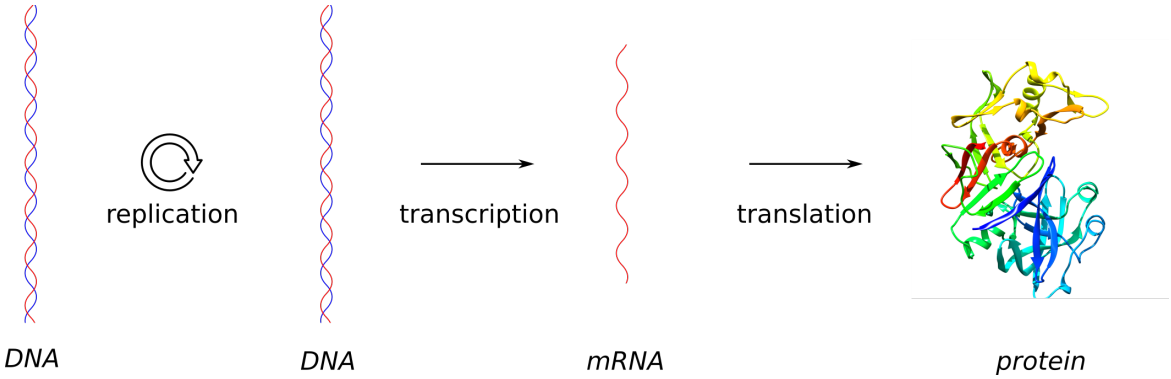



**Fig. 1.1 The central dogma of molecular biology and genetics.** Genetic information contained in DNA is transcribed into messenger RNA (mRNA). mRNA is then transported from the nucleus into the cytoplasm, where the information is translated to protein. Proteins act as the main workhorses of many cellular activities. Replication is the process of doubling the genetic information in DNA prior to inheritance by offspring cells. The illustrated protein is pepsin (PDB: 5PEP), one of the first protein enzymes to be crystallised (Cooper et al., 1990).

et al., 2014; Sclafani and Holzen, 2007). Therefore, there is a growing interest and need to understand in greater detail the mechanistic insight of the replication system.

Fundamentally, DNA replication in prokaryotes, archaea, and eukaryotes is similar (Bleichert et al., 2017; Jacob et al., 1963). In all these domains, DNA is replicated in a semi-conservative manner, which means that new DNA duplex is a combination of a parental template and a complementary newly synthesised strand. Other basic features are also similar, such as replisome function and the 5′ to 3′ directionality of DNA synthesis. In eukaryotes, there are more layers of regulation and pathway redundancy that can partly be attributed to different genomic size, morphology, and complexity (Zhang and O'Donnell, 2016).

The experimental results presented in this dissertation concern the human DNA replication. However, our understanding of replication has come mostly from studies in model organisms such as *Saccharomyces cerevisiae* (budding yeast). Therefore, this introduction will summarise our knowledge of eukaryotic DNA replication by using examples from the *S. cerevisiae* system. Some known species-specific differences occur and are highlighted throughout the dissertation. The cell cycle control of replication is discussed in section 1.2. The three stages of replication initiation, elongation, and termination are reviewed in section 1.3 to 1.5. Lastly, the outstanding questions in the field of eukaryotic DNA replication, some of which this dissertation seeks to address, are introduced in section 1.6.

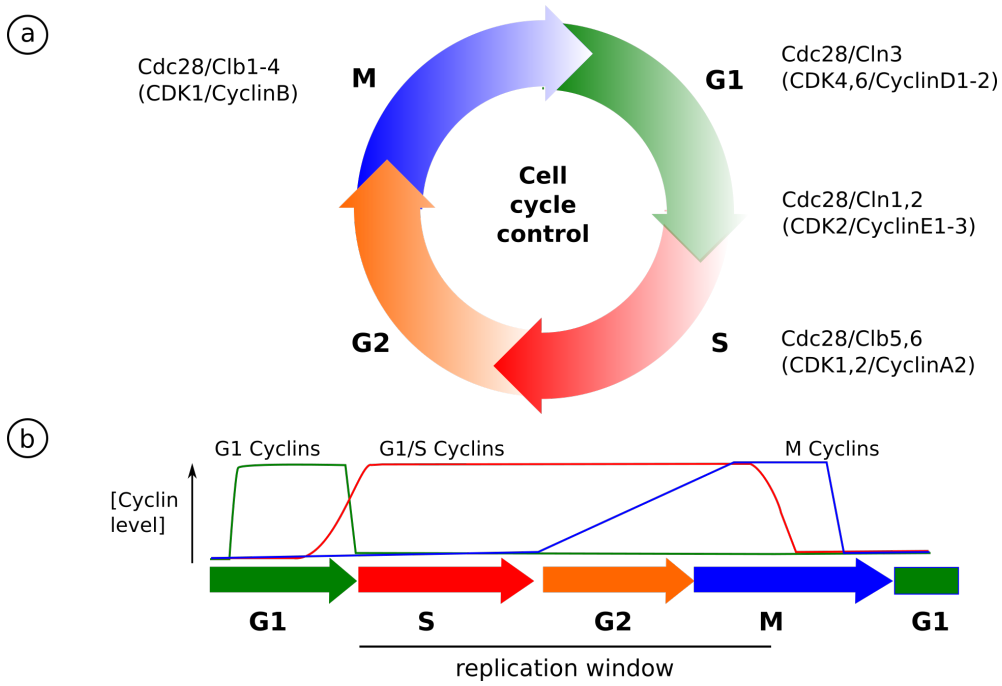



**Fig. 1.2 Cell cycle control. (a)** Eukaryotic cell cycle control and its protein components in *Saccharomyces cerevisiae*. The homologues in higher eukaryotes are described inside parentheses. **(b)** The level of Cyclins in the cells during cell cycle. Replication window is defined within the rising level of G1/S Cyclins, where their activities stimulate DNA replication replication origin firing. See section 1.2 for more information. Clb = Cyclin Type B, Cln = Cyclin. Adapted from (Alberts et al., 2014).

## 1.2 DNA replication and the cell cycle

Any biochemical event that contributes to a successful cell division — such as chromosome condensation, spindle assembly, or DNA replication — is under strict temporal and spatial control of cell cycle. Cell cycle is an orderly sequence of steps leading to a cell doubling its chromosomal contents and dividing into two identical daughter cells. Eukaryotic cell cycle can be divided into four main phases: Gap 1 (G1 phase), Synthesis (S phase), Gap 2 (G2 phase), and Mitotic (M phase) (Alberts et al., 2014).

Cell cycle control relies on the action of Cyclin-Dependent Kinases (CDKs), which consists of two components: the kinase and the activator called Cyclins. In *S. cerevisiae*, only one CDK exists whereas in metazoan there are at least four CDKs (Alberts et al., 2014) (figure 1.2). The concentration of different types of Cyclins oscillates differently during the cell cycle while CDKs remain constant. This cyclical nature of CDK–Cyclin activity defines distinct stages of cell cycle. CDK-Cyclins regulate the progress of several processes by phosphorylating their various protein targets, rendering them active or inactive. Protein targets that are specific for replication are discussed further in the next sections.

DNA synthesis occurs during S-phase, however, there are additional steps that need to be completed beforehand and afterwards. During G1-phase, protein components are assembled on the DNA replication origin. During S-phase, checkpoint mechanism ensures that every part of the genome has been duplicated to avoid missing the information for the progeny. The S-phase Cyclin A switches on the genome duplication process by activating CDK1/2 to phosphorylate several protein targets to begin and to complete DNA replication (a period called 'replication window', figure 1.2b). However, Cyclin A level decreases steeply during the M-phase, which leads to a rapid switching off of the replication programme. This ensures that whole genome duplication occurs *only* during the 'replication window' period and *only* once per cell cycle. Failure to adhere to these strict requirements results in disastrous consequences, such as accumulation of mutations, polyploidy, aneuploidy, and gross chromosomal rearrangements (Hartwell, 1992).

Apart from Cyclin activity, there are other layers of regulation of CDK activity. First is the antagonistic effect of Wee1 kinase and Cdc25 protein phosphatase. Wee1 phosphorylates CDK to render it inactive whereas Cdc25 dephosphorylates it to achieve the opposite effect. Next, the binding of CDK inhibitor proteins (CKIs) such as p27 CKI causes inactivation by inducing structural rearrangement on CDK2–CyclinA (Pavletich, 1999). Lastly, CDK–Cyclins are regulated by the actions of ubiquitin-ligase proteins. Anaphase Promoting Complex or Cyclosome (APC/C) targets specific S- and M-phases Cyclins for degradation (Primorac and Musacchio, 2013). Skp-, Cullin-, F-box-containing complex (SCF) ubiquitin ligase degrades G1/S-phases Cyclins. Intriguingly, SCF can also stimulate CDK activity by degrading certain CKIs (Alberts et al., 2014; Cardozo and Pagano, 2004). In the following sections, the intricate controls that regulate DNA replication are described.

## 1.3 Initiation of DNA replication

The initiation step of replication can be generally divided into three stages: DNA replication origin recognition, licensing and firing. The replication origin recognition step serves to mark locations within the genome where replication may begin. Licensing ensures that once the cell has gone past a certain checkpoint, it cannot go back in cell cycle to re-initiate DNA replication process. This ensures that DNA replication only happens once per cell cycle. Finally, firing is the recruitment step of DNA replication process on the DNA replication origin that carries out the genome duplication. This firing step corresponds to the assembly of two active replicative DNA helicases (termed CMG complex, Cdc45-Mcm2–7-GINS) at each replication origin, that unwind the double-stranded DNA (dsDNA) and expose single-

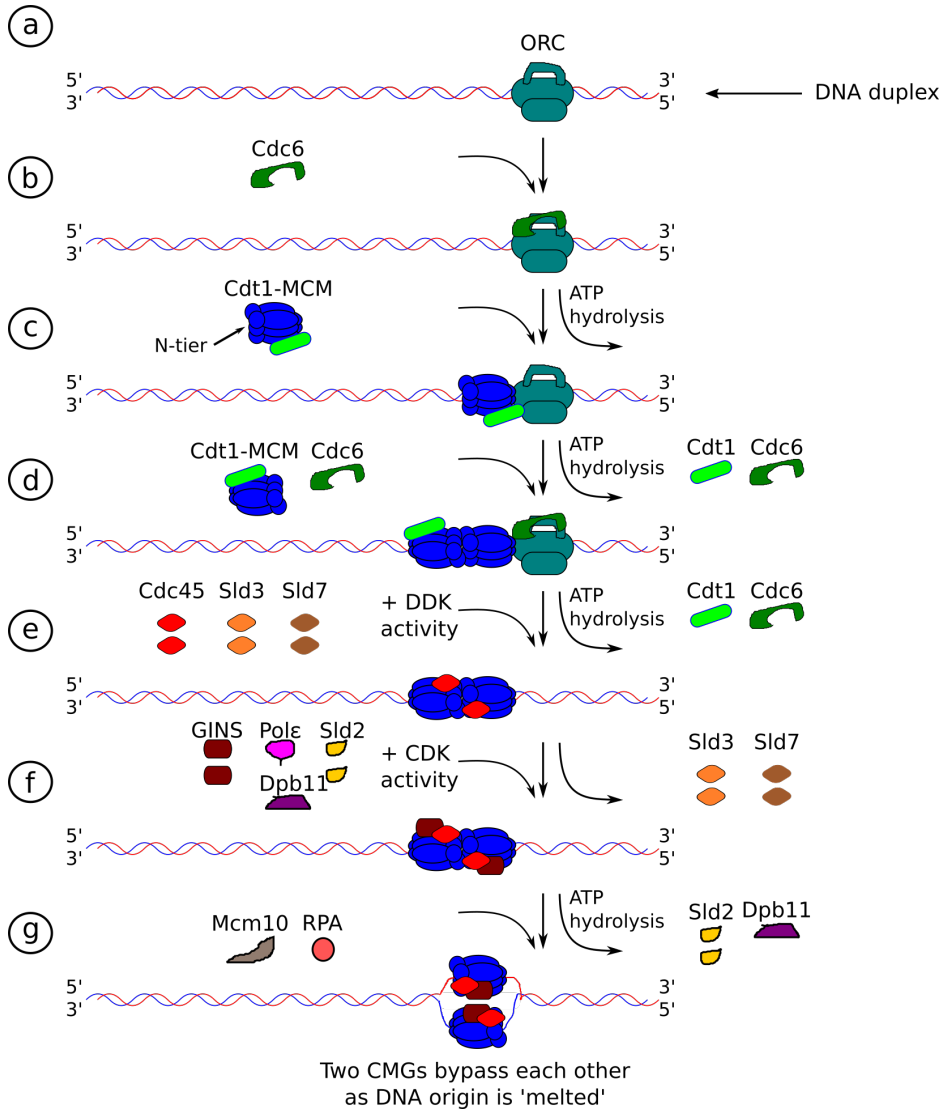



**Fig. 1.3 DNA replication initiation.** Stepwise reactions during DNA replication initiation in the model organism *Saccharomyces cerevisiae* (budding yeast), leading to the formation of active replicative DNA replicative, called the CMG (Cdc45-Mcm2–7-GINS) complex. For step **(e–f)**, ORC is omitted from the diagram for clarity. For step **(g)**, Pol $\varepsilon$ is omitted for clarity. See section 1.3 for more details.

stranded DNA (ssDNA) templates in preparation for nucleotide polymerisation by replicative DNA polymerases (figure 1.3).

### 1.3.1 ORC binding to DNA marks replication origin recognition

The DNA replication origin is a specific location in the genome where the DNA replication programme can begin. A bacterial cell contains only one DNA replication origin within its genome, which has a well-defined sequence called OriC. In a unicellular eukaryote *S. cerevisiae*, there are more than 600 DNA replication origins as demonstrated by comparative genomics with a defined 100- to 150-base pair consensus sequence termed Autonomously Replicating Sequence (ARS) (Siow et al., 2012; Stinchcomb et al., 1979).

In higher eukaryotes, especially human, locating DNA replication origins is not straightforward (Remus et al., 2004; Vashee et al., 2003). There are multitude of differences in terms of genomic size and complexity. The human genome has more DNA replication origins, possibly between 30,000 to 50,000 (Mechali, 2010). Approximately half of these DNA replication origins are located within CpG islands, stretches of DNA where the frequency of the CG di-nucleotide is higher than other regions in the genome (Cadoret et al., 2008). Lastly, access to DNA replication origin may be influenced by transcriptional regulatory elements, transcriptional status, nucleosomal positioning, and epigenetic imprinting (Cadoret et al., 2008; Remus et al., 2004; Vashee et al., 2003).

A DNA sequence that contains multiple guanine repeats may form a specific secondary structure called G-quadruplex (G4) (Prioleau, 2017). A genome-wide mapping study revealed that 70–90% of metazoan DNA replication origins are preceded by potentially G4-forming sequences, collectively called as Origin G-rich Repeated Element (OGRE) (Cayrou et al., 2012). Recent evidences also showed that G4 plays a role in the initiation of DNA replication (Prorok et al., 2019; Valton et al., 2014). However, its exact mechanism is still unclear. It is possible that G4 recruits DNA replication proteins (such as DNA helicase or activation factors), as these proteins contain domains that bind directly to G4. When G4 is formed, the complementary DNA becomes exposed as single-stranded DNA, which may help for direct priming by primase enzyme (priming of replication is discussed further in section 1.4.2). Alternatively, G4 may promote nucleosomal positioning which in turn promotes DNA replication initiation (Bryan, 2019; Hoshina et al., 2013; Kumagai and Dunphy, 2017).

Marking of DNA replication origin is mediated by the Origin Recognition Complex (ORC), a ring structure of six protein subunits that encircle the dsDNA (figure 1.3a) (Bell and Stillman, 1992). Five of the subunits, Orc1–5, are distinct but related in AAA+ ATPase family whereas Orc6 binds to the other subunits through its C-terminal region (Bleichert et al., 2017; Li and Stillman, 2012). The encircling of dsDNA by ORC is an ATP-dependent

process. However, only Orc1 subunit still retains the ATPase activity and is also predicted to be important for the DNA binding activity. Orc4 and Orc5 can bind to but not hydrolyse ATP. Orc2 and Orc3 seem to lose its ability to bind to ATP entirely (Chesnokov et al., 2001; Giordano-Coltart et al., 2005; Klemm et al., 1997).

Structure of *S. cerevisiae*'s ORC bound to dsDNA containing the ACS (ARS-Consensus Sequence) and B1 element (a sequence that is a part of DNA replication origin but distinct from the ACS) offers a snapshot on how the replication origin recognition takes place (Li et al., 2018). ORC binding triggers bending, but not melting, of the dsDNA around the ACS (figure 1.4a). Orc2, Orc5, and Orc6 act cooperatively to bind to the B1 element and to bend the DNA. Viewing from the N-tier and in clockwise manner, Orc1–5 oligomerise around the ACS with arrangement of Orc1-Orc4-Orc5-Orc3-Orc2 (figure 1.4b). A gate exists between Orc1 and Orc2, which possibly serves as an entry for dsDNA. Once the dsDNA enters, Orc2-WHD (Winged-Helix Domain) acts as a temporary lock before Cdc6 seals the gap (compare figure 1.4b and d). Cdc6 is another protein in complex with ORC that is needed to mark the DNA replication origin. DNA duplex forms interactions with the Initiation-Specific Motif (ISM) of all ORC subunits (except Orc4 and Orc6) and the $\beta$-hairpins within Orc2, Orc4, and Orc5. Specific to *S. cerevisiae* system, the sequence specificity of ORC may be explained by the interaction of DNA duplex with the basic path in the Orc1, Orc2, and Orc4. The latter is not present in the higher eukaryote's Orc4 sequence (Kawakami et al., 2015; Li et al., 2018; Yuan et al., 2017).

ORC interacts with DNA replication origin throughout the cell cycle, but this does not mean that replication occurs throughout the cell cycle. Together with another protein factor called Cdc6, they locate the DNA replication origin and act as protein scaffold to deposit DNA replicative helicase onto DNA replication origin (Duzdevich et al., 2015; Mizushima et al., 2000). This subsequent step, called replication licensing, is controlled by cell cycle checkpoints to take place only during G1-phase.

### 1.3.2 Mcm2–7 double hexamer assembly is a hallmark of replication licensing

#### Loading the Mcm2–7 — from ORC to OCCM

Replication licensing is a G1 phase-specific event, in which all the necessary protein components are recruited to the chromatin to begin DNA replication. It was first noticed in *S. cerevisiae*'s DNA replication origin footprint studies (Diffley et al., 1994) and later described as a replicative helicase loading step (Labib et al., 2001). Limiting helicase recruitment in

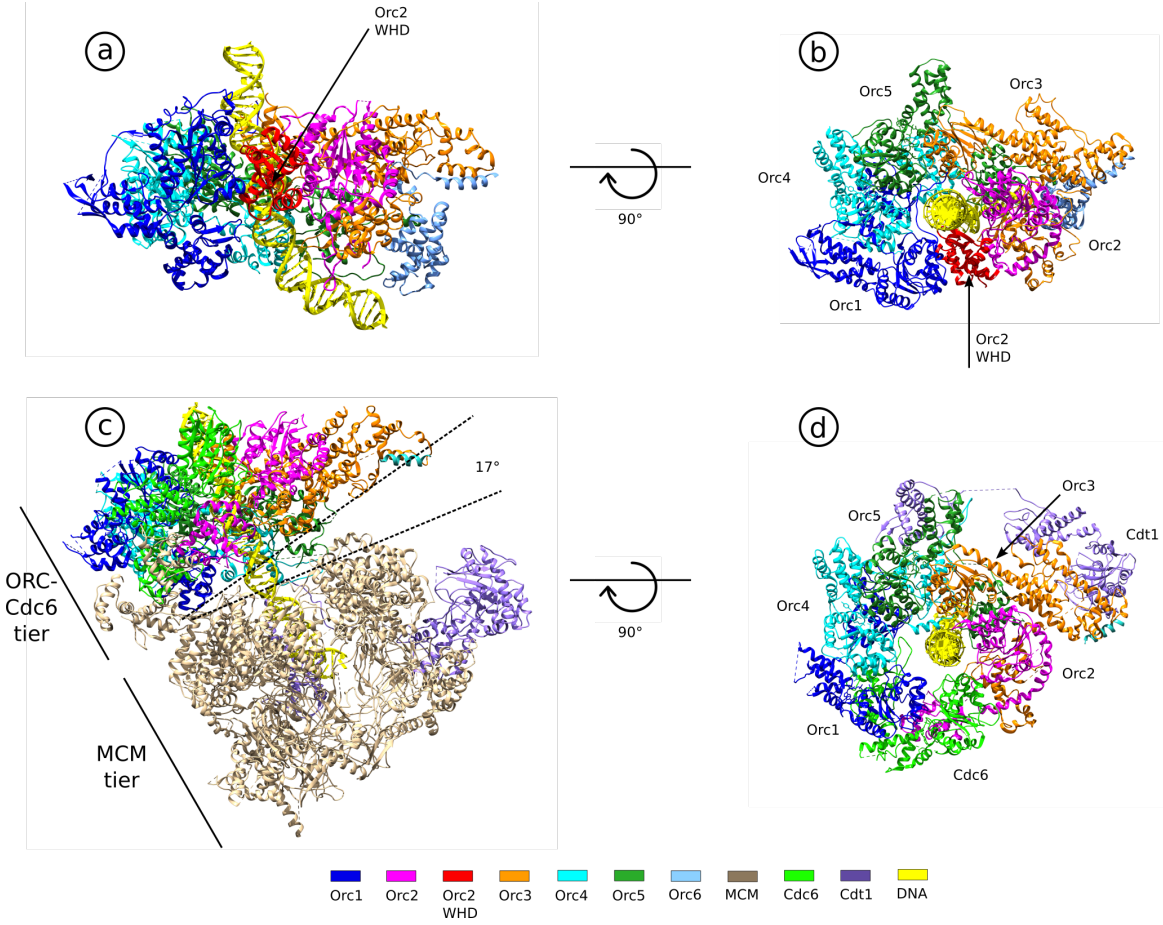



**Fig. 1.4 Structure comparison of ORC in isolation or in the context of OCCM (ORC-Cdc6-Cdt1-Mcm2–7) bound to DNA. (a)** Structure of *S. cerevisiae*'s ORC+DNA (PDB: 5ZR1). **(b)** Top view of ORC+DNA, showing the arrangement of ORC around DNA. The Winged Helix Domain (WHD) of Orc2 in (a) and (b) is coloured in red. **(c)** Structure of *S. cerevisiae*'s OCCM+DNA (PDB: 5UDB). **(d)** Top view from the N-tier of ORC of OCCM+DNA. Mcm2–7 subunits are omitted for clarity. ORC = Origin Recognition Complex, OCCM = ORC-Cdc6-Cdt1-Mcm2–7. (Li et al., 2018; Yuan et al., 2017).

the end of G1-phase is one of the regulatory steps to ensure that replication only happens once per cell cycle (Siddiqui et al., 2013).

The core of eukaryotic replicative DNA helicase is a ring of six MCM (Mini Chromosome Maintenance) proteins. MCMs were first identified in *S. cerevisiae*'s genetic screens as replication initiation mutants (Dutta and Bell, 1997; Gibson et al., 1990; Maine et al., 1984). The MCM ring is made up of six non-identical subunits, Mcm2–7, belonging to the same AAA+ (ATPases Associated with diverse cellular Activities) protein family (Vijayraghavan and Schwacha, 2012). This is in contrast with replicative DNA helicase in both prokaryotic and archaeal systems, where it comprises six identical subunits (Brewster and Chen, 2010). Each of the eukaryotic Mcm2–7's subunit is indispensable for cellular growth, indicating unique role for each subunit during replication (Labib et al., 2000; Yan et al., 1991). Multiple studies demonstrated that Mcm2–7 is the main component of the eukaryotic replicative helicase (Aparicio et al., 1997; Dutta and Bell, 1997; Labib et al., 2000). Curiously, Mcm2–7 was shown to possess a very weak DNA unwinding activity *in vitro* (Bochman and Schwacha, 2008; Davey et al., 2003) that suggests Mcm2–7 requires other protein co-activators for more robust helicase activity (section 1.3.3).

Mcm2–7 recruitment to the DNA replication origin is mediated by a chaperone called Cdt1 (Caillat and Perrakis, 2012). Cdt1 is a regulatory target to control the Mcm2–7 loading mechanism strictly during G1 phase (Caillat and Perrakis, 2012; Truong and Wu, 2011). The Cdt1–Mcm2–7 complex exists as a heptamer in an ATP-dependent manner (Zhai et al., 2017b). It appeared that Mcm2–7 — in complex with Cdt1 — adopts a 'lock-washer' conformation with an open gate between Mcm2 and Mcm5 subunit (Zhai et al., 2017a). This Mcm2–Mcm5's open gate is also conserved in *Drosophila melanogaster* (fruit fly) (Costa et al., 2011, 2014). However, the width of this gate (10–15 Å) is too small for dsDNA entry (diameter 25–30 Å). This was supported by biochemical experiments showing that Mcm2–7 by itself could not encircle dsDNA and the inability to purify Mcm2–Mcm5 dimer (Bochman et al., 2008; Bochman and Schwacha, 2007; Bowers et al., 2004; Davey et al., 2003). C-terminal extension (CTE) of Mcm5 is located inside Mcm2–7's central channel, which also helps to prevent unregulated or accidental insertion of dsDNA (Zhai et al., 2017b). Nevertheless, the gate between Mcm2 and Mcm5 must be the entry point of dsDNA as linking these two subunits (but not between other Mcm2–7 subunits) abrogates Mcm2–7 loading to the DNA (Samel et al., 2014).

The loading of Mcm2–7 helicase to the DNA replication origin cannot be executed by Cdt1–Mcm2–7 alone. Instead, it is intricately regulated by cell control mechanism (G1-CDK activity), interaction with ORC–Cdc6 complex, and orderly ATP binding and hydrolysis (Bowers et al., 2004; Coster et al., 2014). G1-CDK phosphorylates DNA-bound

Orc2 and Orc6 subunits, which triggers binding to Cdc6 (figure 1.3b) (Chen and Bell, 2011). In addition, it was demonstrated that S-phase CDK has an influence on inhibition of pre-replication complex formation (Chen and Bell, 2011). Cdc6 is required to fully 'lock' the DNA duplex inside the ring formed by Orc1–5 and Cdc6 (compare figure 1.4b and d) (Yuan et al., 2017). In an ATP-dependent reaction, ORC–Cdc6 complex subsequently recruits the Cdt1–Mcm2–7 complex, mediated and regulated by the C-terminal region of Mcm3, resulting in the formation of a transient but stable helicase loading complex of ORC-Cdc6-Cdt1-Mcm2–7 (OCCM) complex (figure 1.3c) (Evrin et al., 2014; Fernandez-Cid et al., 2013; Frigola et al., 2013; Randell et al., 2006; Remus et al., 2009). Orderly and ATP-dependent releases of Cdc6, Cdt1, and possibly ORC from the OCCM complex results in the deposition of single copy of Mcm2–7 hexamer on the DNA replication origin (Evrin et al., 2009; Remus et al., 2009; Ticau et al., 2015).

DNA-bound OCCM's high-resolution structure was described (Yuan et al., 2017). The overall architecture of the complex is a two-ring structure, 17° tilt with respect to each other, and trapped DNA duplex within the central channel of both rings. One ring is made up of Orc1–5 subunits and Cdc6 and the other consists of Mcm2–7 hexamer. The C-terminal of Mcm2–7 hexamer makes contact with the C-tier of ORC–Cdc6 ring. Cdt1 forms multiple interactions with Mcm2–7 subunits of Mcm2, Mcm4, and Mcm6 (figure 1.4c).

The structure of DNA-bound OCCM provides a further insight on how DNA is being inserted through the OCCM's central channel. From the ORC–Cdc6 ring, DNA interacts with the $\alpha$-helix in Orc4 and positively-charged loops of Orc2 and Orc6 (Yuan et al., 2017). DNA also forms multiple contacts with the Mcm2–7 ring via helix-2-insert (H2I) of Mcm2, Mcm4, Mcm6 and presensor 1 (PS1) of Mcm4, Mcm6, Mcm7 (Yuan et al., 2017).

Mcm2–7 has to adopt a gate that is wide enough for dsDNA insertion, which means widening the gap between Mcm2 and Mcm5. The structure of DNA-bound OCCM showed that the electron density for the Mcm5 is less clearly defined, which suggests that Mcm5 is flexible with respect to other subunits and likely to facilitate the wider opening of the Mcm2–Mcm5 gate (Yuan et al., 2017). ORC and Cdc6 must also induce conformational change in Cdt1–Mcm2–7 complex from 'lock-washer' conformation to more planar, although the details of this step remain unknown. The OCCM complex also helps to orient the Mcm2–7 ring to align with and to insert the dsDNA, by inducing the conformational change of Mcm5. From the OCCM structure, a general arrangement of the Mcm2–7 ring can be defined: seen from C-tier in clockwise direction, Mcm2–7 is arranged as Mcm5-Mcm3-Mcm7-Mcm4-Mcm6-Mcm2 (figure 1.5a).

### Loading of Mcm2–7 double hexamer (MCM-DH)

Since replication occurs in bi-directional manner, two copies of the Mcm2–7 need to be loaded on DNA replication origin in an opposing orientation. Two Mcm2–7 rings are loaded sequentially on DNA in a head-to-head double hexamer conformational shape during G1-phase and this process requires ATP hydrolysis (Evrin et al., 2009; Remus et al., 2009). The loading process is dependent on ordered association and dissociation of different Cdc6 and Cdt1 molecules via ATP hydrolysis by Cdc6 (figure 1.3d) (Ticau et al., 2015). However, how many copies of ORC are required remains open to discussion. There are two proposed mechanisms: one-ORC versus two-ORC models. The earlier was supported by single-molecule study that showed only one ORC was needed as the second Mcm2–7 loading was supported through interaction with the first loaded Mcm2–7 complex (Ticau et al., 2015). This was also backed by electron microscopy study that described one copy of ORC is sufficient to bind to Mcm2–7 double hexamer (Sun et al., 2014). The two-ORC model was substantiated by a biochemical experiment that used Mcm3 C-terminal mutant. This particular Mcm3 mutant has lost its ability to interact with ORC–Cdc6 (i.e. a poor substrate for first loaded Mcm2–7), but it should still be a functional substrate for the second Mcm2–7 recruitment as postulated by one-ORC model. However, the loading of second Mcm2–7 ring harbouring Mcm3 mutant was not observed (Frigola et al., 2013).

Structures of Mcm2–7 double hexamer (MCM-DH) alone or bound to dsDNA showed that it is arranged in a head-to-head conformation, with a 14° wedge between them (figure 1.5b) (Abid Ali et al., 2017; Li et al., 2015; Noguchi et al., 2017). MCM-DH assembly is mediated by inter-hexameric interactions between the N-terminal domains: zinc-fingers (ZFs) in the B-subdomain, N-terminal extension (NTE) of Mcm3, Mcm5, Mcm7, and N-terminal insertion (NTI) of Mcm7 (Li et al., 2015). MCM-DH structure with dsDNA revealed nucleotide occupancy, likely to be ADP as a byproduct of ATP hydrolysis, between the Mcm5–Mcm2, Mcm2–Mcm6, Mcm7–Mcm3, Mcm5–Mcm2 (Noguchi et al., 2017). The structure shows that dsDNA is threaded throughout the entire channel of Mcm2–7 ring, which suggests that MCM-DH does not perform the DNA unwinding (figure 1.5b). This is in agreement with previous biochemical observations that Mcm2–7 requires co-activators to perform DNA unwinding (Bochman and Schwacha, 2008; Davey et al., 2003). Both Mcm2–7 hexamers interact with the dsDNA in spiral-like manner, mediated by the Presensor 1 (PS1) loops of all Mcm2–7 subunits except Mcm2 (Noguchi et al., 2017).

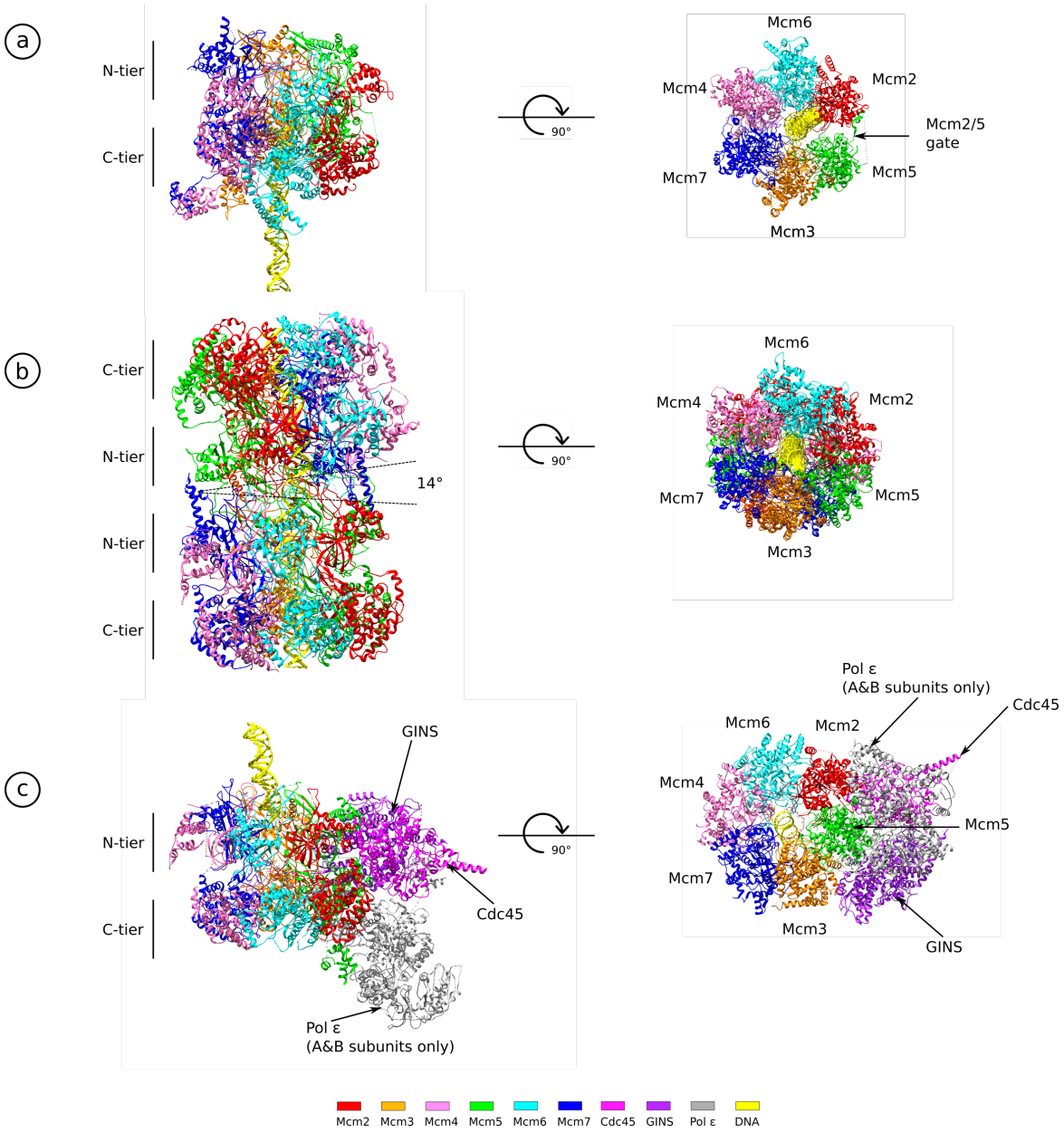



**Fig. 1.5 Structure comparison of Mcm2–7 in the context of OCCM, MCM-DH, and CMG–Pol$\varepsilon$.** **(a)** Structure of *S. cerevisiae*'s Mcm2–7 as a part of OCCM complex (PDB: 5UDB). The gate between Mcm2–Mcm5 is indicated with an arrow. ORC, Cdc6, and Cdt1 are omitted for clarity. **(b)** Structure of *S. cerevisiae*'s Mcm2–7 double hexamer (MCM-DH) bound to dsDNA (PDB: 5BK4). The angle of the wedge between two Mcm2–7 hexamers is indicated. The gate between Mcm2-Mcm5 is closed and the Mcm2–7 subunits arrangement is more compact. **(c)** Structure of *S. cerevisiae*'s Mcm2–7 as a part of CMG complex with Pol$\varepsilon$ and DNA substrate (PDB: 6HV9). Positions of Cdc45, GINS, and Pol$\varepsilon$ are indicated with their respective arrows. For simplicity, all four subunits of GINS are grouped as one colour. Mcm2–7 = Mini Chromosome Maintenance 2–7, OCCM = ORC-Cdc6-Cdt1-Mcm2–7, MCM-DH = Mcm2–7 double hexamer, CMG–Pol$\varepsilon$ = Cdc45-Mcm2–7-GINS and Pol $\varepsilon$. (Goswami et al., 2018; Noguchi et al., 2017; Yuan et al., 2017).

### Control of replication origin licensing

The replication licensing step needs to be perfectly controlled in order to prevent re-replication in S-phase. There are some regulatory principles that are shared across the eukaryotes. Cell-cycle control through a rising level of G1/S-CDK phosphorylates Cdc6, which is a signal for ubiquitin-dependent degradation of Cdc6 (Drury et al., 2000; Mimura et al., 2004). Cdt1–Mcm2–7 binding with the ORC is also disrupted by S-phase CDK (S-CDK) phosphorylation on Orc2 and Orc6 (Chen and Bell, 2011; Fernandez-Cid et al., 2013; Frigola et al., 2013). Steric hindrance mediated by RXL-motif in N-terminal of Orc6 also prevented effective Cdt1–Mcm2–7 interaction (Nguyen et al., 2001; Wilmes et al., 2004). Down-regulation of Cdc6 transcription and nuclear export of the Cdt1–Mcm2–7 complex also provide extra layers of regulation of replication origin licensing (Bell and Labib, 2016).

There are regulatory features that are specific to higher eukaryotes. Firstly, ubiquitination of Orc1 by $SCF^{Skp2}$ acts as a signal for proteolysis (DePamphilis et al., 2006; Siddiqui et al., 2013). Secondly, Cdc6 is also degraded by proteasome, in which the signalling process is regulated by protein phosphatase PP2A (Kalfalah et al., 2015). Next, ubiquitination of chromatin-bound Cdt1 by PCNA-dependent Cul4–$Ddb1^{Scf2}$ and S-CDK-dependent $SCF^{Skp2}$ prevent replication origin licensing (Arias and Walter, 2007; Truong and Wu, 2011). Lastly, Geminin (that is absent in *S. cerevisiae*, but present in related fungi *S. pombe*) inhibits replication by binding to Cdt1 which leads to impairment of Mcm2–7 loading, but not of ORC chromatin binding (Wohlschlegel et al., 2000; Wu et al., 2014). Geminin is absent during G1-phase, which allows replication origin licensing to occur. Its level builds up during S- and G2-phase before being ubiquitinated and degraded at M-phase (McGarry and Kirschner, 1998).

MCM-DH binding to the DNA is a hallmark of eukaryotic replication. However, at this stage DNA unwinding has not taken place yet. The replication origin has to be 'fired' to begin DNA unwinding and hence DNA synthesis. This replication origin firing step is a highly complex and regulated step, which is discussed in the following section.

## 1.3.3 CMG helicase activation is a signature of DNA replication origin firing

Although Mcm2–7 is the main catalytic core of replicative DNA helicase, in itself it has very weak DNA unwinding capability (Bochman and Schwacha, 2008; Davey et al., 2003). It requires two co-activators — Cdc45 (Cell Division Cycle 45) and GINS (Go-Ichi-Ni-San, Japanese for 5-1-2-3 that corresponds to its subunits Sld5-Psf1-Psf2-Psf3) — for more robust unwinding ability. Both Cdc45 and GINS maintain interaction with Mcm2–7 throughout

DNA replication progression (Pacek and Walter, 2004; Takayama et al., 2003; Tercero et al., 2000). The isolated CMG complex — produced either endogenously or recombinantly — was shown to possess efficient DNA unwinding activity *in vitro* (Gambus et al., 2006; Ilves et al., 2010; Moyer et al., 2006). The highly regulated processes of assembly and activation of CMG complex can be considered as the culmination of replication origin firing (Bell and Labib, 2016).

### S-CDK and DDK control the DNA replication origin firing

Origin firing is regulated by inter-dependent actions of two highly-conserved kinases: S-phase CDK (S-CDK) and Dbf4-Dependent Cdc7 Kinase (DDK) (Zegerman, 2015). S-CDK consists of Cdc28 kinase and their regulatory subunits Clb5/Clb6 (figure 1.2). Meanwhile, DDK comprises Cdc7 kinase and Dbf4 regulatory subunit. These kinases ensure timely and proper recruitment of Cdc45 and GINS to the DNA-bound Mcm2–7 by phosphorylating various key targets prior to replication origin firing (Heller et al., 2011; Labib, 2010; Yabuuchi et al., 2006; Yeeles et al., 2015).

DDK activity, but not necessarily S-CDK, drives the recruitment of Cdc45 to the DNA-bound Mcm2–7 (figure 1.3e). Currently, the only known target for DDK during replication is Mcm2–7 — most likely the N-terminal tail of Mcm2, Mcm4 and Mcm6 (Cho et al., 2006; Francis et al., 2009; Hardy et al., 1997; Sheu and Stillman, 2010). Deletion of Mcm4's N-terminal domain bypasses the DDK requirement, which suggests a mechanism where DDK phosphorylation suppresses the inhibitory action of Mcm4 N-terminal domain (Sheu and Stillman, 2010) and possibly exposes the binding site for protein Sld3 (Treslin is higher eukaryotes) (Abid Ali et al., 2017; Deegan et al., 2016). Sld3 was also shown to be a chaperone protein that recruits Cdc45 (discussed in greater detail in section 1.6.2 and 1.6.3) (Kamimura et al., 2001). *In vivo* experiments demonstrated that Sld3 and Cdc45 are inter-dependent for binding to Mcm2–7, although *in vitro* experiments suggested that Sld3 could directly bind to the Mcm2–7 without Cdc45 (section 1.6.4) (Deegan et al., 2016; Kamimura et al., 2001; Tanaka et al., 2011a). Sld3 is stabilised by binding of another protein Sld7 (MTBP in higher eukaryotes), although Sld7 is not essential for replication (Boos et al., 2013; Tanaka et al., 2011a; Yeeles et al., 2015). The interplay between Sld3/Treslin and Cdc45 is discussed further in section 1.6.

Meanwhile, S-CDK has only two essential targets: Sld2 (RecQL4 in higher eukaryotes) and Sld3 (Tanaka et al., 2007; Zegerman and Diffley, 2007). Phosphorylation of Sld2 and Sld3 create phospho-binding sites for N-terminal BRCT repeats in Dpb11. This interaction is also conserved in higher eukaryotes (Boos et al., 2011). This S-CDK-dependent reaction triggers the formation of a transient pre-loading complex (pre-LC) that comprises Sld2, Sld3,

Dpb11, GINS, and DNA polymerase $\varepsilon$ (Pol$\varepsilon$). All the components of pre-LC are needed to complete the CMG helicase formation by recruiting the final components, GINS (figure 1.3f) (Muramatsu et al., 2010). The requirement of Pol$\varepsilon$ even before DNA synthesis is intriguing, but it is confirmed by both biochemical and structural studies that showed only one of its subunits, Dpb2, is essential for GINS recruitment (Goswami et al., 2018; Sengupta et al., 2013).

### Architecture and directionality of CMG helicase

The structure of the CMG helicase bound to forked DNA (DNA oligomer mimicking the replication fork) showed that the N-tier of Mcm2–7 is rigid in context of Cdc45 and GINS associations (figure 1.5c) (Georgescu et al., 2017; Goswami et al., 2018). The C-tier of CMG is very flexible and adopts two major conformations of 'extended' and 'compact' shape, perhaps capturing two different states of CMG activity during DNA unwinding. It is likely that the N-tier of CMG is required for processivity while the C-tier is the main motor for DNA translocation — similar to what was observed in archaeal model (Brewster and Chen, 2010; Georgescu et al., 2017). Furthermore, the CMG structure also suggests a mechanism for Cdc45's role during DNA unwinding. Cdc45 sits on top of the Mcm2–Mcm5 'gate' interface, where it is poised to prevent ssDNA release from the 'gate' and to act as a 'wedge' during DNA unwinding event (Costa et al., 2014; Petojevic et al., 2015; Szambowska et al., 2014).

The structure of CMG–forked DNA also presented evidence that ssDNA, instead of dsDNA, is present throughout the central channel of CMG complex (figure 1.5c) (Georgescu et al., 2017; Goswami et al., 2018). This proves that CMG unwinds the dsDNA by steric-exclusion mechanism, in which the lagging-strand is completely excluded from the N-tier opening. This structural study is in agreement with DNA unwinding assay performed using purified CMG complex with DNA substrates that contain non- or covalently attached lagging-strand blocks (Abid Ali et al., 2016; Fu et al., 2011; Kose et al., 2019; Yuan et al., 2016). Moreover, this structure established the directionality of CMG translocation, in which the complex travels N-face first from $3'$ to $5'$ (Georgescu et al., 2017; Goswami et al., 2018).

Based on the observations that one CMG complex contains only one Mcm2–7 ring and encircles ssDNA throughout, there must be a major structural rearrangement step to convert DNA-bound MCM-DH to two separate copies of CMG complex that bypasses each other in the DNA replication origin (figure 1.3f and g). At our current knowledge, the details of this conversion step remain unclear. DNA topology-based experiment using purified *S. cerevisiae* proteins suggested that orderly substitution from Mcm2–7-bound ADP to the ATP is caused by the structural arrangement of Mcm2–7 due to the association with other

protein co-activators. Furthermore, association of Mcm10 protein stimulate initial DNA melting (Douglas et al., 2018). However, this step has not been analysed by structural or single-molecule approaches.

#### Temporal control of DNA replication origin firing

It needs to be highlighted that not all licensed replication origins are fired. Furthermore, not all replication origins are fired at the same time. The reason for different replication timing programmes (early- vs late-firing replication origins) is still unclear (Bell and Labib, 2016). It may be due to intrinsic difference in the DNA sequence, for example ARS1 DNA element in *S. cerevisiae* is early-firing, whereas ARS501 is late-firing (Ferguson and Fangman, 1992). Meanwhile, DNA replication replication origins in the centromere are always early-firing whereas replication origins in the telomere fire late (Raghuraman et al., 2001). This difference in DNA sequence may result in varying efficiency of helicase loading. Nucleosomal rearrangement may also influence the timing, as shown by the effect of Rpd3 histone deacetylase that delays firing (Knott et al., 2009; Vogelauer et al., 2002). These varying replication programmes and dormant replication origins may offer a fail-safe mechanism. When early-fired replisome encounters replication fork block or collapse, the late-firing and dormant replication origins can be fired as 'back-up' to ensure complete genomic duplication within the replication window.

Once the replication origin is fired and CMG helicase is active, a specific DNA structure called replication fork is formed. It is a region on DNA where DNA synthesis takes place, as it exposes a pair of ssDNA templates as substrates for replicative DNA polymerases. DNA polymerases then begin template-directed nucleotide polymerisation during the progression of of DNA replication.

## 1.4 Progression of DNA replication

### 1.4.1 Replisome Progression Complex (RPC)

The replisome consists of several enzymes such as helicase, polymerase, primase, and topoisomerase. *In vitro* experiments showed that each of these enzymes can function independently from each other (Bell and Labib, 2016). However, *in vivo* conditions require replisome components to be closely assembled and regulated. This is to ensure tight coupling of DNA unwinding and synthesis activity to minimise exposure of single-stranded (ssDNA) from genomic stress surveillance system. In addition, replisome assembly helps to

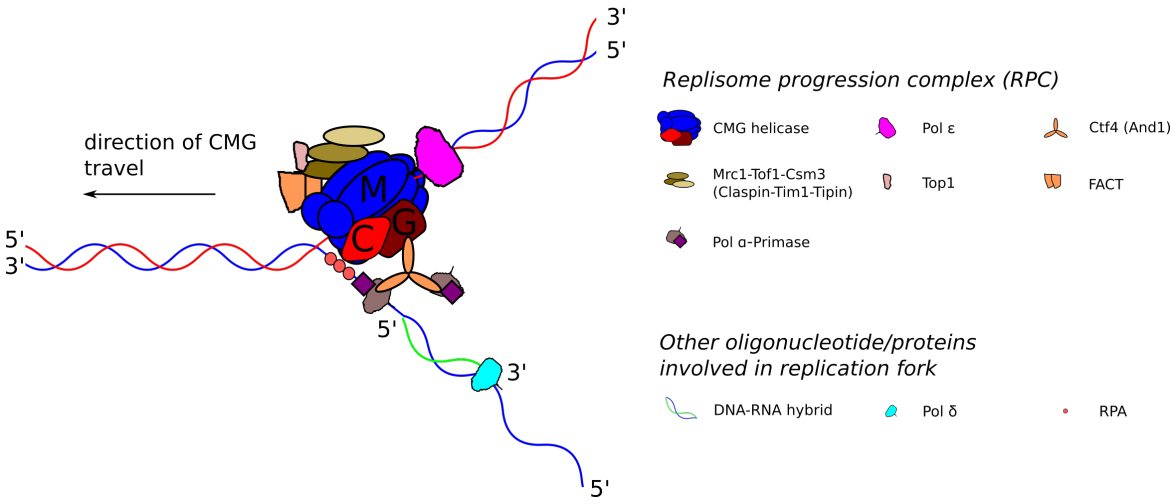



**Fig. 1.6 Replisome Progression Complex (RPC) in a replication fork.** Proteins that made up Replisome Progression Complex (RPC) in replication fork to ensure normal progression of DNA replication. Red = leading-strand DNA, Blue = lagging-strand DNA. See section 1.4 for more details.

deal with challenging situations such as nucleosome remodelling, replication obstacles, and coordination with checkpoint systems (Bell and Labib, 2016; Kim et al., 1996).

In a co-purification experiment coupled with mass spectrometry analysis, a multi-protein assembly that associated with replication fork was identified. This complex, referred to as Replisome Progression Complex (RPC), consists of a CMG helicase core with Ctf4 adaptor protein (And1 in human), Top1 isomerase that removes positive DNA supercoils, FACT histone chaperone, Mrc1 checkpoint mediator (Claspin), Tof1 (Timeless), and Csm3 (Tipin) (Gambus et al., 2006). Later, the RPC was also confirmed to interact with two replicative DNA polymerases, Pol$\alpha$ and Pol$\varepsilon$ (Gambus et al., 2009; Sengupta et al., 2013).

Mcm10 was also found to co-purify with RPC in high-salt condition (Gambus et al., 2006). Mcm10's role during replication is still largely unclear. Functional cell assays that deplete Mcm10 level showed that it is required for initial DNA unwinding activity by the CMG, complete replisome assembly, replication fork progression, and DNA repair (Alver et al., 2014; Chattopadhyay and Bielinsky, 2007; Douglas et al., 2018; Park et al., 2008). Mcm10 was also found to associate with replisome components of the CMG helicase, Pol$\alpha$ and PCNA (Das-Bradoo et al., 2006; Mayle et al., 2019; Warren et al., 2008). During elongation step, Mcm10 is also thought to preserve CMG interaction with DNA and subsequent replisome stability on the fork (Wasserman et al., 2018). The structure of Mcm10 in complex with CMG or other replisome components has not been reported yet. However, data from chemical cross-linking coupled with mass spectrometry of purified CMG–Mcm10 complex suggests that Mcm10 is likely to be located in the N-face of the CMG complex (Mayle et al., 2019).

### 1.4.2 Priming of DNA synthesis

DNA synthesis is carried out by three highly-conserved, multi-subunit B-family DNA polymerases: Pol$\alpha$, Pol$\delta$, and Pol$\varepsilon$. All three polymerases share common traits, in which they require free 3′-OH (hydroxyl) group prior to DNA synthesis and polymerise only in 5′ to 3′ direction (Johansson and Dixon, 2013).

The oligonucleotide primers that provide the hydroxyl group are generated by a highly conserved DNA-dependent RNA polymerase called Primase complex, which consists of Pri1 (PriS in human) catalytic subunit and Pri2 (PriL in human) regulatory subunit (Pellegrini, 2012). The iron-sulphur domain that exists within the PriL was shown to be essential for priming in eukaryotes (Klinge et al., 2007). The PriS catalytic subunit of Pol$\alpha$ can be further divided into five domains: inactive exonuclease domain (in other DNA polymerases, this exonuclease is active and required for proofreading, which suggests why Pol$\alpha$ is selectively disadvantaged for bulk DNA synthesis), N-terminal domain, 'fingers', 'palm', and 'thumb'. The 'palm' possesses primer-extension activity, the 'finger' engages with the incoming nucleotides, and the 'thumb' helps to position the RNA–DNA hybrid primer with the DNA template (Pellegrini, 2012).

Primase also forms a specific and constitutive complex with Pol$\alpha$ and its B subunit, mediated by the C-terminus of the catalytic subunit of Pol$\alpha$ (Muzi-Falconi et al., 2003). This hetero-tetrameric complex initiates the DNA synthesis by synthesising RNA–DNA hybrid primer: 7–12 nucleotides (nt) of RNA by Primase, followed by 10–15 nt DNA extension by Pol$\alpha$. Pol$\alpha$ specifically recognises an A-form helix, formed by DNA template and RNA primer, and then begins synthesising DNA oligonucleotides to extend the RNA primer (Perera et al., 2013).

In order to couple DNA unwinding and priming, Primase–Pol$\alpha$ complex forms an indirect contact with CMG helicase via the adaptor protein called Ctf4 (Gambus et al., 2009; Tanaka et al., 2009; Zhu et al., 2007). Trimeric Ctf4 was shown to associate simultaneously with a single CMG helicase and two Pol$\alpha$, mediated by three identical binding sites for a specific motif called Ctf4-Interacting Peptide (CIP) found in both Sld5 (a GINS subunit) and Pol$\alpha$ (Simon et al., 2014). This observation suggests a model for lagging strand synthesis, in which Ctf4 serves as a protein hub for repeated priming of lagging strand, similar to bacterial replisome that contains two lagging-strand polymerases (Reyes-Lamothe et al., 2010; Simon et al., 2014). However, the role of Ctf4 is not limited to bridge CMG with Primase–Pol$\alpha$, as it was shown to mediate CMG interactions with other CIP-containing factors, such as Tof2 and Dna2 (Villa et al., 2016). Specific to the human system, And1 was demonstrated to interact with Mcm10 and Pol$\alpha$ via a unique C-terminal HMG (High Mobility Group) domain (Kilkenny et al., 2017; Zhu et al., 2007).

### 1.4.3 Leading and lagging strand synthesis

Since each fired replication origin produces bi-directional replication and DNA polymerisation can only occur in 5′ to 3′ direction, the replisome has to accommodate continuous leading strand synthesis (i.e. same direction as CMG helicase translocation) and discontinuous lagging strand synthesis. In addition, the replisome has to include DNA polymerases that are highly processive and contain proofreading mechanism to prevent unwanted error during DNA synthesis. It has been generally accepted that Pol$\varepsilon$ carries out the bulk leading strand synthesis from a single priming event, whereas the Pol$\delta$ performs the lagging strand synthesis from multiple priming events, producing short DNA stretches called Okazaki fragments (McElhinny et al., 2008; Pursell et al., 2007; Sakabe and Okazaki, 1966). Chromatin-Immunoprecipitation (ChIP) analysis combined with strand-specific sequencing found that Pol$\varepsilon$ preferentially binds to leading strand whereas the Pol$\delta$ to the lagging strand, providing the basis for division of labour between the polymerases (Yu et al., 2014). However, detailed mechanisms on how the polymerases prefer their respective strand remain unknown.

Pol$\varepsilon$ is a large, multi-subunit, and highly-conserved enzyme that possesses two catalytic activities: template-directed DNA synthesis and 3′ to 5′ exonuclease proofreading ability (Hogg and Johansson, 2012). Its role is not limited to DNA replication, as there are evidences of Pol$\varepsilon$'s involvement in epigenetic silencing, sister chromatid cohesion, and DNA repair (Pursell and Kunkel, 2008). Pol$\varepsilon$ consists of four subunits: catalytic Pol2 (PolE in human) for both DNA synthesis and exonuclease activity, regulatory subunit Dpb2 (PolE2 in human) that is essential for cell viability, and Dpb3 and Dpb4 small subunits (p12 and PolE3 in human, respectively) (Hogg and Johansson, 2012). Intriguingly, N-terminal catalytic domain of Pol$\varepsilon$ is not essential for cell survival, although replication may be slowed down (Dua et al., 1999; Kesti et al., 1999) — highlighting functional redundancy in the DNA synthesis. Instead, the regulatory Dpb2 subunit of Pol$\varepsilon$ is vital for cell viability, as it is required for binding to the CMG and subsequently stimulating the DNA unwinding activity (Georgescu et al., 2014; Langston et al., 2014; Sengupta et al., 2013). These biochemical observations were supported by structure of CMG–Pol$\varepsilon$, in which the C-terminal half of Dpb2 interacts with CMG helicase via Mcm2, Mcm6, and Cdc45 (Goswami et al., 2018; Sun et al., 2015).

Pol$\delta$ shares similar characteristics with Pol$\varepsilon$, where its catalytic subunit (A subunit) contains both polymerase and 3′ to 5′ exonuclease proofreading activities (Tahirov, 2012). The other regulatory subunits are required to form interactions with the catalytic subunits and other processivity factor called PCNA (Tahirov, 2012). There is no report of direct tethering of Pol$\delta$ with RPC, which may suggest its activity is uncoupled from CMG activity (De Piccoli et al., 2012; Sengupta et al., 2013). However, the repeated priming by the Primase–Pol$\alpha$ complex in the lagging strand template allows Pol$\delta$ to synthesise numerous Okazaki

fragments which will later be joined as one contiguous DNA during Okazaki fragment maturation process to complete DNA synthesis (Bell and Labib, 2016; Sakabe and Okazaki, 1966).

The processivity of both polymerases is improved by a trimeric Proliferating Cell Nuclear Antigen (PCNA) clamp. The DNA polymerisation rate is increased by two-fold for CMG-bound Pol$\varepsilon$ and by at least 100-fold for Pol$\delta$ (Chilkova et al., 2007; Georgescu et al., 2014). The loading of PCNA to the DNA is mediated by a pentameric clamp loader called Replication Factor C (RFC) that temporarily disassembles and then rapidly loads PCNA on the DNA. Interestingly, RFC preferably recruits PCNA on the primer–template junction, which suggest a mechanism of handing over of newly-synthesised DNA from Primase–Pol$\alpha$ to Pol$\delta$ and/or Okazaki fragment maturation factors (Yao and O'Donnell, 2012).

The facts that the catalytic domain of Pol$\varepsilon$ is dispensable for cell viability and Pol$\delta$ (together with the Primase–Pol$\alpha$) can still perform the leading-strand synthesis are intriguing. Experiments using mutant Pol$\varepsilon$ of *S. cerevisiae* — where the mutant is catalytically inactive but still retains the ability to interact with CMG — showed interdependent actions of two replisomes at the beginning of DNA synthesis, in which Pol$\delta$ synthesises DNA from a preceding Okazaki fragment and subsequently 'catches up' to and hands over the newly synthesised DNA to the Pol$\varepsilon$ in the leading strand replisome (Aria and Yeeles, 2018). This finding is surprising as it rules out the necessity for leading strand priming, further highlighting the complexity and redundancy in the system for division of labour during DNA synthesis.

Although the error rate is very low, DNA polymerases may incorporate the wrong mismatched base or damaged nucleotide to the newly-synthesised DNA strand (Cortez, 2019). To resolve this issue, DNA replication is also coupled with DNA damage mechanisms of mismatch repair (MMR) and base excision repair (BER). Human MMR proteins, mainly MutS$\alpha$ and MutL$\alpha$, are linked with PCNA which may also explain why MMR is more prevalent in the lagging strand (Hombauer et al., 2011; Kleczkowska et al., 2001). BER protein component of RnaseH2 also interacts with PCNA, mediated by its PCNA-interacting protein (PIP) (Bubeck et al., 2011). The resulting excised base is subsequently resolved by Okazaki fragment maturation process, which is discussed further in section 1.4.4 (Dungrawala et al., 2015).

To complete DNA synthesis, RNA primers produced by Primase need to be removed. In the case of leading strand, Pol$\varepsilon$ can remove it through its exonuclease activity. Meanwhile, the situation for Pol$\delta$ is rather more complicated as it requires Okazaki fragments maturation machinery.

### 1.4.4 Okazaki fragments maturation

In a normal replicating mammalian cell, there are about 50 million Okazaki fragments that need to be processed properly to ensure a successful cell division (Zheng and Shen, 2011). This processing mechanism is termed Okazaki fragments maturation, which involves four steps: strand displacement DNA synthesis by Pol$\delta$, formation of DNA flap, processing of the flap by specific endonucleases, and finally ligation of the nicked DNA (figure 1.7) (Zheng and Shen, 2011).

Unlike other replicative polymerases, Pol$\delta$ is unusual in a way that it can 'invade' dsDNA to continue DNA synthesis in a process known as strand displacement (figure 1.7a). The dsDNA structure displaced by Pol$\delta$ is the preceding primer annealed to the template. As a result, a 5′-unannealed ssDNA flap that contains the RNA–DNA primer is generated (figure 1.7b) (Stodola and Burgers, 2016; Zheng and Shen, 2011).

This short flap structure — that can be either single or double flap — is then resolved by Fen1 (Flap Endonuclease 1) (Liu et al., 2004; Tsutakawa et al., 2011). Fen1 recruitment to the flap is mediated by PCNA, via its PIP motif (figure 1.7c) (Gary et al., 1999). In some cases of longer DNA flap, Dna2 endonuclease is more preferred than Fen1 (Ayyagari et al., 2003). Finally, the nicks generated by both endonucleases are ligated by DNA Ligase I, which is also recruited by PCNA (figure 1.7d), to finally join all the Okazaki fragments to form a continuous DNA (figure 1.7e) (Turchi et al., 1994).

These Okazaki fragment maturation factors (Pol$\delta$, PCNA, RFC, Fen1, and DNA Ligase I) are also required during DNA repair process, specifically during long-patch base excision repair where DNA flap is also generated after the action of Glycosidase and Ape1 proteins. This shows that both processes of DNA replication and repair are intertwined with each other, as the protein machinery is partially shared (Zhou and Elledge, 2000). In case Okazaki fragment maturation fails and generates unligated fragments, a back-up pathway involving PARP1 is activated to resolve the unligated fragments via single-strand break repair machinery (Hanzlikova et al., 2018).

### 1.4.5 Nucleosome rearrangement during replication

DNA inside our cell is incredibly compact. The approximate length of all our chromosomal DNA joined together is about 2 metres, but it needs to be packaged carefully to fit inside a nucleus that is 6 $\mu$m in diameter (Alberts et al., 2014). DNA has to be assembled into a macromolecular structure called chromatin. The building block of chromatin is nucleosome, that consists of 146 base pairs of DNA in a left-handed helical turn around histone core particle (Alberts et al., 2014). The core histones form an octamer, comprising two dimers

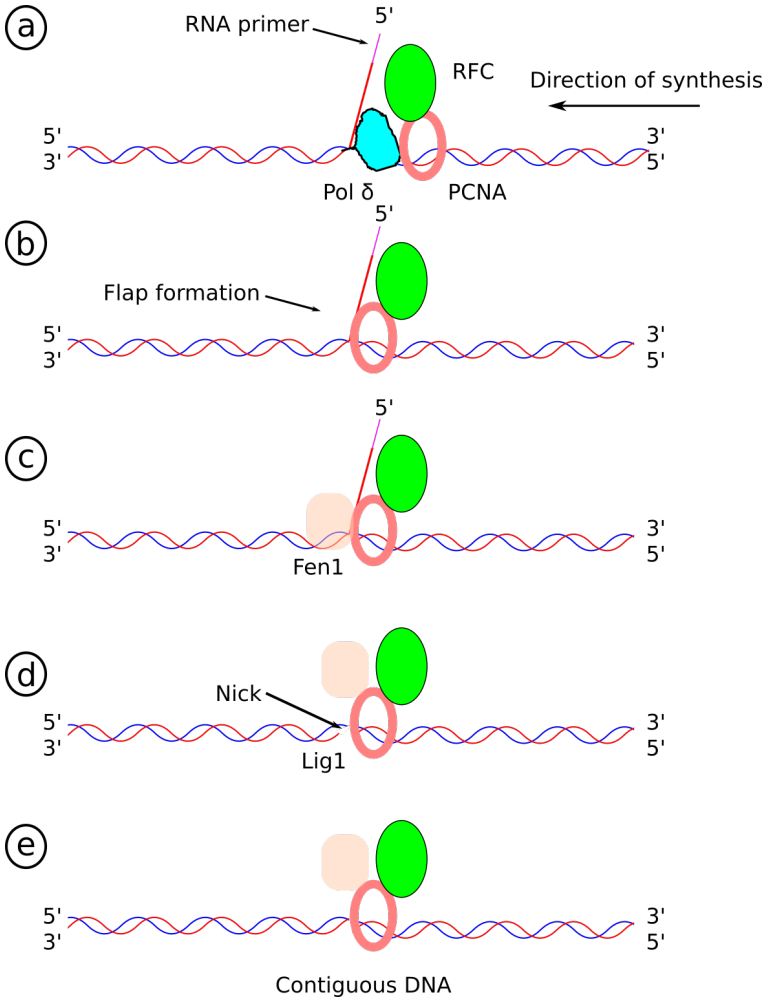



**Fig. 1.7 Okazaki fragments maturation.** Steps for maturation of Okazaki fragments. **(a)** Strand displacement by Pol$\delta$. **(b)** Formation of DNA flap and dissociation of Pol$\delta$. **(c)** Fen1 recruitment to process the flap. **(d)** Recruitment of DNA Ligase 1 (Lig1) to process the nick created by Fen1. **(e)** Contiguous dsDNA is produced after Okazaki fragment maturation step. Adapted from (Bell and Labib, 2016).

H2A–H2B and tetrameric H3–H4. In addition, histone core particles are subjected to multiple post-translational modifications, such as methylation and acetylation, which forms the molecular basis of epigenetic information (Whitehouse and Smith, 2013).

Therefore, in the context of whole chromosome replication, replisome encounters two additional challenges. Firstly, it has to efficiently disrupt the nucleosome ahead of the replication fork to gain access for DNA template. Secondly, it needs to reassemble the nucleosome immediately behind the replication fork to keep all the epigenetic information intact and packaging of DNA tight (Whitehouse and Smith, 2013).

Several replisome components possess histone-binding activity, which may play a major role in disrupting the nucleosome and transferring the parental H3–H4 dimers. The N-terminal tail of Mcm2 contains a conserved tyrosine motif that binds to the parental histones (Foltman et al., 2013). Mutational studies and the crystal structure of Mcm2 bound to histone showed that these conserved residues are the major contact points of the complex. Interestingly, Mcm2's N-terminal tail may also mimic the DNA for binding to the outside residues of H3–H4 (Huang et al., 2015; Richet et al., 2015).

Pol$\alpha$ was shown to interact with the H2A–H2B complex, via its histone-binding motif similar to Mcm2 and mediated by Ctf4 (Evrin et al., 2018). In addition, small subunits of Pol$\varepsilon$, Dpb3–Dpb4 complex (PolE3–PolE4 in human), are able to interact directly with H3–H4 (Bellelli et al., 2018). Lastly, histone chaperone FACT (Facilitates Chromatin Transaction) complex was also shown to interact with the replisome progression complex and helped to destabilise nucleosome structure (Chen et al., 2018). FACT consists of two subunits, SPT16 that binds to H2A–H2B and SSRP1 that interacts with H3–H4 (Chen et al., 2018).

After nucleosomal disassembly and DNA synthesis, nucleosomes need to be re-assembled tightly and immediately to preserve epigenetic information (Lucchini and Sogo, 1995). A burst of histone production during S-phase is also required to maintain the nucleosome density post replication (De Koning et al., 2007). Depositing tetrameric H3–H4 — that contains most of the epigenetic data — into the new nucleosome is likely to be mediated by histone chaperones such as Asf1 and Caf1 (Katan-Khaykovich and Struhl, 2011; Masumoto et al., 2005; Schneider et al., 2006). The acetylated H3–H4 (in H3 lysine 56) is passed on from Asf1 to Caf1, which in turn is positioned around the replication fork by PCNA (Driscoll et al., 2007).

### 1.4.6 Dealing with replication blocks

The replisome has to ensure that it copies billions of DNA base pairs faithfully and accurately. Along the way, it encounters a number of challenging regions such as DNA lesions, secondary structures adopted by DNA repeats such as G-quadruplex (GQ), and 'difficult-to-replicate'

loci such as fragile sites. Failure to deal with these frequent obstacles can result in base-pair mutations, gross chromosomal rearrangements, and chromatin segregation difficulties. These can potentially result in human medical disorders such as higher predisposition to cancer, neurodegenerative diseases, and early ageing (Barlow and Nussenzweig, 2014; Bicknell et al., 2011; Hills and Diffley, 2014; Jackson et al., 2014; Sclafani and Holzen, 2007).

Interference with normal replisome progression can lead to replication fork stalling, which has to be resolved to prevent genomic instability (Errico and Costanzo, 2012). This section discusses briefly on how the replisome deals with these problems.

### Helicase and polymerase uncoupling

Replication fork stalling is an event when DNA unwinding is uncoupled from DNA synthesis. This occurs when the CMG helicase unwinds past DNA damage but DNA polymerases fail to synthesise DNA. This event exposes ssDNA that is rapidly coated with Replication Protein A (RPA) (figure 1.8) (Zou and Elledge, 2003). One way to deal with this problem and to preserve the integrity of stalled replication fork is to pause replication. During pausing, replisome has to continuously associate with the replication fork to ensure a speedy resumption of replication after DNA repair is completed. In addition, reloading of the replisome component to the DNA is not possible outside G1-phase (Calzada et al., 2005; Errico and Costanzo, 2012).

A conserved group of proteins that is likely to modulate and bridge helicase and polymerase activity is called replication pausing complex (figure 1.8). It consists of four proteins: Mrc1, Tof1, Csm3, and Ctf4 (Claspin, Tim1, Tipin, and And1 in human, respectively) (Errico and Costanzo, 2012). These four proteins were shown to interact with CMG in the replication fork (Gambus et al., 2006; Osborn and Elledge, 2003). Depletion of replication pausing complex was shown to slow down fork progression rate by 50–70%, although the exact mechanism remains to be understood (Hodgson et al., 2007; Szyjka et al., 2005; Tourriere et al., 2005). Reconstitution studies using purified human proteins showed that Tim1–Tipin directly interacts with CMG helicase components and decreases the rate of DNA unwinding (Cho et al., 2013). Meanwhile, Tim1–Tipin interaction with the DNA polymerases can be detected via co-immmunoprecipitation but not with purified proteins. This suggested a possibility in which Tim1–Tipin interaction with the DNA polymerases *in vivo* is mediated by adaptor protein And1 (Cho et al., 2013; Errico et al., 2009). Nevertheless, depletion of this complex, especially Tim1–Tipin, results in the accumulation ssDNA in the cell that is likely due to persistent uncoupling of helicase and polymerase activity (Errico et al., 2007; Smith et al., 2009).

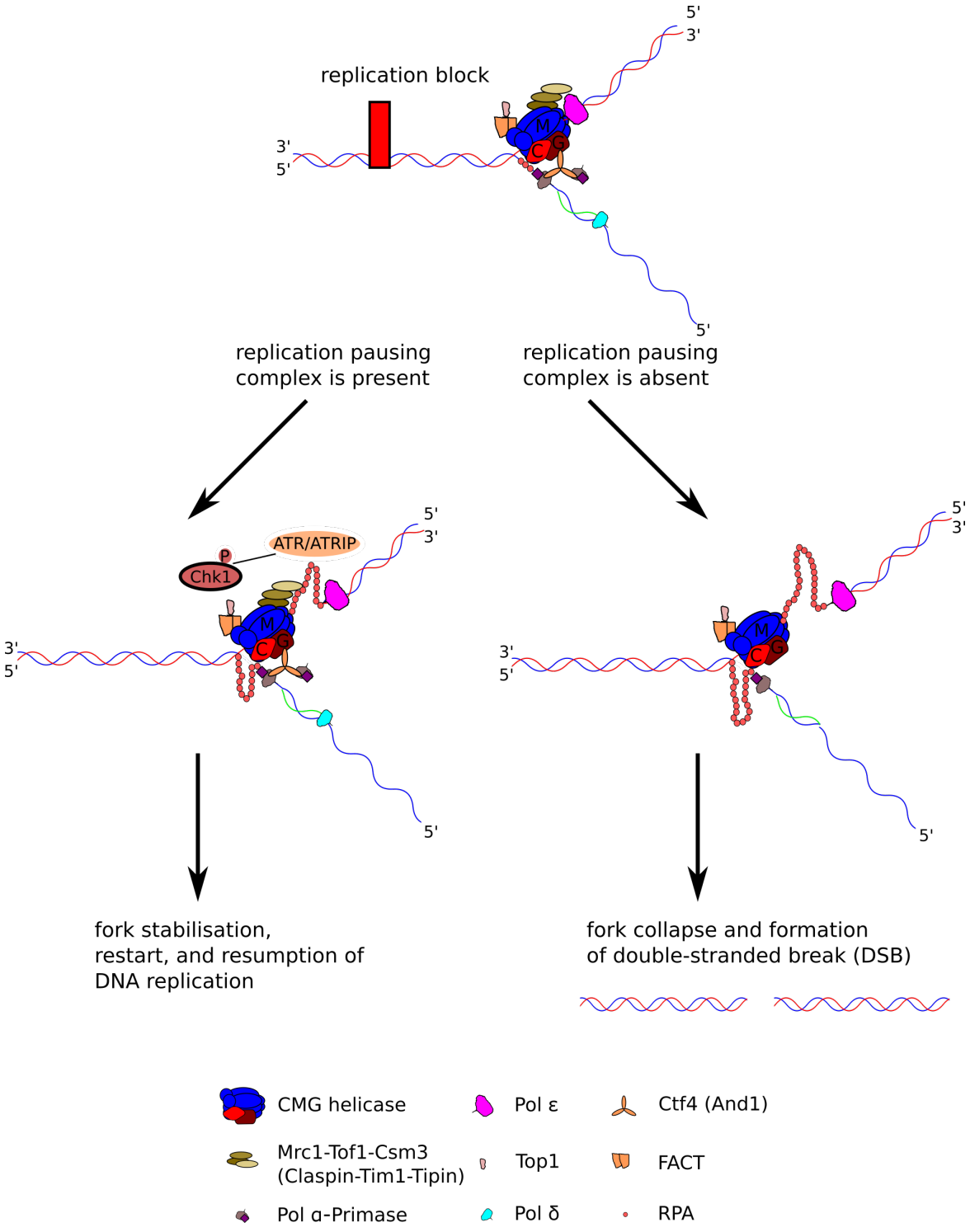



**Fig. 1.8 Replication pausing complex.** The replication pausing complex, that consists of Mrc1-Tof1-Csm3-Ctf4 (Claspin-TimI-Tipin-And1 in human), helps to maintain the stability of replication fork upon replication block. As the replication stalls, Replication Protein A (RPA) quickly coat single-stranded DNA produced from CMG activity. Replication pausing complex helps to recruit the ATR/ATRIP which phosphorylates effector kinase protein, Chk1. In the absence of replication pausing complex, the replication fork collapses and double stranded break (DSB) is formed. See section 1.4.6 for more information. Adapted from (Errico and Costanzo, 2012).

Replisome pausing complex also plays a major role in stabilisation of the replication fork upon stalling by ensuring the activation of so-called 'intra-S phase' checkpoint control (figure 1.8) (Errico and Costanzo, 2012). The intra-S phase checkpoint is a highly regulated surveillance system that specifically safeguards the replication fork integrity and prevents disassembly of replisome components (Nyberg et al., 2002). The main sensor of this checkpoint system is ATR kinase, whereas the effector kinase is Chk1 (Rad53 in *S.cerevisiae*) (Branzei and Foiani, 2006; Nyberg et al., 2002). ATR's recruitment to the replication fork and its auto-phosphorylation is stimulated by RPA-coated ssDNA and mediated by ATRIP (ATR-Interacting Protein) (Rouse and Jackson, 2002; Zou and Elledge, 2003). ATR activity is further enhanced by TopBP1, Rad9-Rad1-Hus1 (9-1-1 complex), and Rad17/RFC (Errico and Costanzo, 2012). Fully active ATR phosphorylates the effector kinase Chk1, which is mediated by replication pausing complex components such as Tof1 and Mrc1 (Naylor et al., 2009; Osborn and Elledge, 2003). This signalling step results in prevention of mitotic entry, inhibition of replication origin firing, stabilisation of stalled forks, and restarting the replication fork after DNA repair has been completed (Errico and Costanzo, 2012).

There are alternative ways to deal with a replication block, which are replication bypass or fork reversal (Cortez, 2019). DNA synthesis can resume after polymerase bypass by rapid re-priming by Primase–Pol$\alpha$ in the lagging strand or by PrimPol in the leading strand. The block is resolved later by translesion bypass polymerase (Bianchi et al., 2013; Garcia-Gomez et al., 2013; Taylor and Yeeles, 2018). Meanwhile, fork reversal occurs when nascent and DNA template strands anneal to form a four-way DNA structure. This event in higher eukaryotes is likely to be dependent on fork-reversal enzymes such as SMARCAL1 and its resolution is likely to involve homologous recombination machinery such as Rad51 and MRN (Mre11-Rad50-Nbs1) complexes (Betous et al., 2012; Dungrawala et al., 2015; Zellweger et al., 2015).

### Helicase blocking

Replication blocks such as inter-strand crosslinks (ICL) or DNA–protein crosslinks (DPC) can interfere with CMG helicase translocation (Cortez, 2019). When these blocks are present in the lagging strand, CMG can bypass them due to its strand-exclusion unwinding mechanism (Fu et al., 2011; Kose et al., 2019; Yuan et al., 2016). However, when the blocks are in the leading strand CMG requires helps from other proteins. The detailed mechanism for ICL and DPC bypass in the leading strand is still largely unknown. Our current knowledge suggests that RTEL1 is needed for bypass of the ICL by providing a ssDNA patch for the CMG to latch onto (Sparks et al., 2019). In the case of the DPC, CMG bypass requires FANC-M, BLM helicase, and ATR kinases (Ling et al., 2016; Rohleder et al., 2016).

Nevertheless, ICL and DPC damaged DNA needs to be repaired to allow normal chromosome segregation. In higher eukaryotes for ICL repair is dependent on FANC-M complexes or TRAIP ubiquitination of CMG subunit that stimulates recruitment of NEIL3 glycosylase (Semlow et al., 2016; Wu et al., 2019). The resulting gap needs to be replicated by TLS polymerase. For DPC, studies in *X. laevis*' egg extracts showed that proteolysis of the crosslinked protein is possible through the action of either SPRTN or proteasomes (Larsen et al., 2019; Stingele et al., 2016).

#### Double-strand breaks (DSBs)

Double-strand breaks (DSBs) is caused by template strand breaks, fork reversal cleavage by nucleases, or endonuclease cleavages which poses a serious impediment for replisome progression (Cortez, 2019). The homologous recombination (HR) machinery is required for DSB repair, especially during replication as homologous template is present (Errico and Costanzo, 2012). There is evidence that the HR machinery plays a direct role during DNA replication to prevent and repair DSB. Upon replication stalling, Mre11-Rad50-Xrs2 (MRN complex Mre11-Rad50-Nbs1 in higher eukaryotes) is recruited to the replication fork to help stabilising the fork. Mre11 depletion results in replication-specific accumulation of DSBs (Costanzo et al., 2001). Recruitment of Mre11 to the replication fork is risky, considering that it contains a nuclease activity that may degrade the nascent ssDNA. It was speculated that Mre11 nuclease activity is required for proper template generation before template switch. Other HR machinery components such as Brca2-dependent Rad51 may also be required to coat the nascent ssDNA, thereby preventing ssDNA from nuclease degradation, stabilising the replication fork, performing strand invasion to initiate HR, and ultimately allowing replication restart (Hashimoto et al., 2010; Schlacher et al., 2011; Tarsounas et al., 2003; Thorslund and West, 2007). The details on how the DSB is linked to replication machinery remain an intense subject of research.

## 1.5 Termination of DNA replication

Unlike other stages of replication, detailed mechanisms of DNA replication termination are much less understood. When two converging replication forks meet, DNA synthesis is always terminated and replisome is rapidly disassembled. Therefore, this step needs to be highly regulated to prevent premature stoppage of replication. The main factors that specify regions of termination are the position and time of initiation between two adjacent replication origin sites. Termination should occur halfway between the two replication origins that are

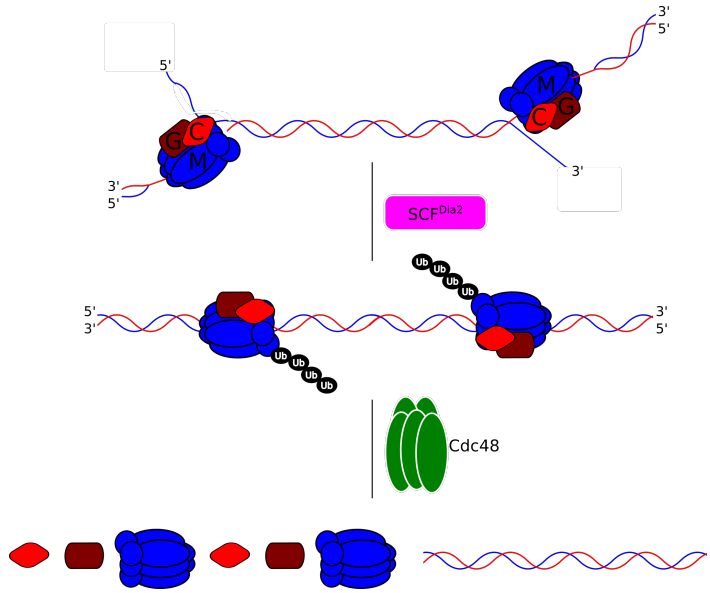



**Fig. 1.9 Termination of DNA replication.** Termination of DNA replication in the eukaryotes, based on *X. laevis*' egg extract system. SCF$^{Dia2}$ an E3 ubiquitin ligase that polyubiquitylates Mcm7. This serves as a signal for CMG degradation by Cdc48 segregase. Ub = ubiquitin. Other components of replisome are omitted for simplicity. See section 1.5 for more details. Adapted from (Bell and Labib, 2016).

fired at the same time, regardless of the obstacles between the two sites (Hawkins et al., 2013; Smith and Whitehouse, 2012).

The CMG complex is likely to be the main target for replisome disassembly and represents the replication termination step (figure 1.9). Ordered removal of the CMG is mediated by an F-box protein called Dia2 and an E3 ubiquitin ligase called SCF$^{Dia2}$ (Maric et al., 2014; Pan et al., 2006). The Mcm7 subunit is the specific target for ubiquitination by SCF$^{Dia2}$ (Maric et al., 2014). This ubiquitination process is enhanced by the interaction of the N-terminal domain of SCF$^{Dia2}$ with the Ctf4 and Mrc1 proteins of the replisome progression complex (Morohashi et al., 2009). Poly-ubiquitination of Mcm7 is likely to be the signal for degradation of the CMG by Cdc48 segregase (Dewar et al., 2015). However, the molecular mechanisms on how Cdc48 segregase specifically recognises ubiquitinated Mcm7 as a part of the terminating CMG helicase and whether this ubiquitinated Mcm7 is really essential for the CMG disassembly remain to be understood. Another model was proposed where the structure of the CMG helicase is altered when two replication forks meet, which may explain the specificity of the disassembly step (Wellinger and Zakian, 2012). A recent *in vitro* study using purified *S. cerevisiae*'s proteins showed that Pif1-family helicases, such

as Pif1 and Rrm3, are required to resolve the torsional stress formed when two replisomes converge (Deegan et al., 2019).

The termination of replication in the telomeric region is not fully understood, yet. In the telomere, only one replication fork reaches the end. The CMG helicase may induce a similar proposed structural change and simply slides off the chromosome (Wellinger and Zakian, 2012). In the case where the CMG has not been disassembled at the end of S-phase, there is an M-phase pathway where Mcm7 is ubiquitinated by TRAIP ubiquitin ligase and subsequently degraded by the VCP segregase (Gambus et al., 2018).

## 1.6 Recruitment of CMG helicase co-activator Cdc45

The recruitment of the CMG helicase co-activator, Cdc45, to the DNA-replication origin bound Mcm2–7 is an essential step for a proper DNA replication initiation in eukaryotes. Genetic screens and biochemical studies using *S. cerevisiae* systems highlighted the importance of a protein called Sld3 for this particular step. Based on limited sequence bioinformatics analysis and several biochemical experiments, the human protein Treslin was identified as the functional orthologue of Sld3. However, the molecular mechanisms of how Treslin participates in the recruitment of Cdc45 to Mcm2–7 remain unclear. The focus of this project is to elucidate the role of Treslin in the assembly and activation of the CMG helicase.

This section summarises our current knowledge about Cdc45, Treslin, and Mcm2–7 during DNA replication initiation. In section 1.6.1, the role of Cdc45 as the main component of the CMG helicase is described. Section 1.6.2 discusses the function of Sld3 in ensuring normal initiation of DNA replication in *S. cerevisiae* and subsequently how Treslin was identified as the functional orthologue of Sld3 in human. Section 1.6.3 and 1.6.4 present the latest knowledge available in literature regarding Treslin's interactions with the main components of CMG helicase, Cdc45 and Mcm2–7.

### 1.6.1 Cdc45 in DNA replication

The *CDC45* (Cell Division Cycle 45) gene was first identified in a *S. cerevisiae* genetic screen for normal progression of the cell cycle (Hennessy et al., 1991). A loss-of-function mutation caused an increased rate of mitotic chromosome loss and recombination. A co-immunoprecipitation study showed that the gene product of *CDC45* is required for DNA replication and that it interacts with at least two MCM subunits (Hopwood and Dalton, 1996; Zou et al., 1997).

A series of experiments showed that Cdc45 is required throughout the replication process. Cdc45 is involved during the initiation step, where its association with the replication origin is dependent on DDK-phosphorylated Mcm2–7 (Mimura et al., 2000; Sheu and Stillman, 2010). Cdc45 is also needed to continually engage with replisome during the fork progression (Gambus et al., 2006; Tercero et al., 2000). Cdc45's exact role was finally identified as a co-activator of replicative DNA helicase (Gambus et al., 2006; Masuda et al., 2003; Pacek et al., 2006; Pacek and Walter, 2004).

Cdc45 is well-conserved across all eukaryotes (figure 1.10a). The human Cdc45 is made up of 566 amino acids with overall molecular mass of 65.6 kDa. The crystal structure of human Cdc45 was described, in which it forms a tight disk-shaped structure and its N-terminal's DHH and C-terminal's DHHA1 domains form an extensive interaction (figure 1.10b) (Simon et al., 2016).

The structure of Cdc45 as a part of CMG complex was also described (figure 1.10c). Here, Cdc45 forms interactions with the N-tier of Mcm2–7 (Abid Ali et al., 2016; Sun et al., 2015). This is mediated by a CMG-interacting domain (CID) that interacts with the A-subdomains of Mcm2 and Mcm5. The point of contact within Cdc45 involves a hydrophobic motif and a conserved phenylalanine in the position 367 (Simon et al., 2016). Meanwhile, Cdc45 also forms an interaction with the B-domain of the Psf1 subunit of GINS via the augmented DHH domain and acidic loop within CID (Abid Ali et al., 2016; Simon et al., 2016). Based on this information, there are two proposals for the molecular role of Cdc45 in the context of CMG helicase. Cdc45 may act as a block to prevent single-stranded DNA escape after unwinding (Costa et al., 2014; Petojevic et al., 2015). Another suggestion is that Cdc45 acts as a wedge that directly splits the parental strand (Szambowska et al., 2014).

Structural comparison between human Cdc45, archaeal GAN (GINS-Associated Nuclease), and bacterial RecJ revealed a surprising evolutionary link between them (figure 1.10d, e, and f) (Pellegrini, 2017). All three proteins possess DHH and DHHA1 domains that are bridged with flexible helical stem (Krastanova et al., 2012; Sanchez-Pulido and Ponting, 2011). This was first identified with bioinformatic analysis of the N-terminal sequence and confirmed by the crystal structures (Cheng et al., 2016; Oyama et al., 2016; Sanchez-Pulido and Ponting, 2011; Simon et al., 2016). Cdc45 and GAN share similar properties as being the replicative DNA helicase co-activator, while RecJ is strictly a prokaryotic DNA repair protein that contains a $5'$ to $3'$ ssDNA exonuclease activity. Interestingly, GAN still retains its exonuclease activity whereby Cdc45 has lost it. This observation presents us with a curious evolutionary link. Eukaryotic Cdc45 has evolved to carry out a specific replication task, losing its exonuclease activity along the way. Meanwhile, archaeal GAN did not lose the exonuclease activity. The crystal structure suggested that the extensive intra-molecular

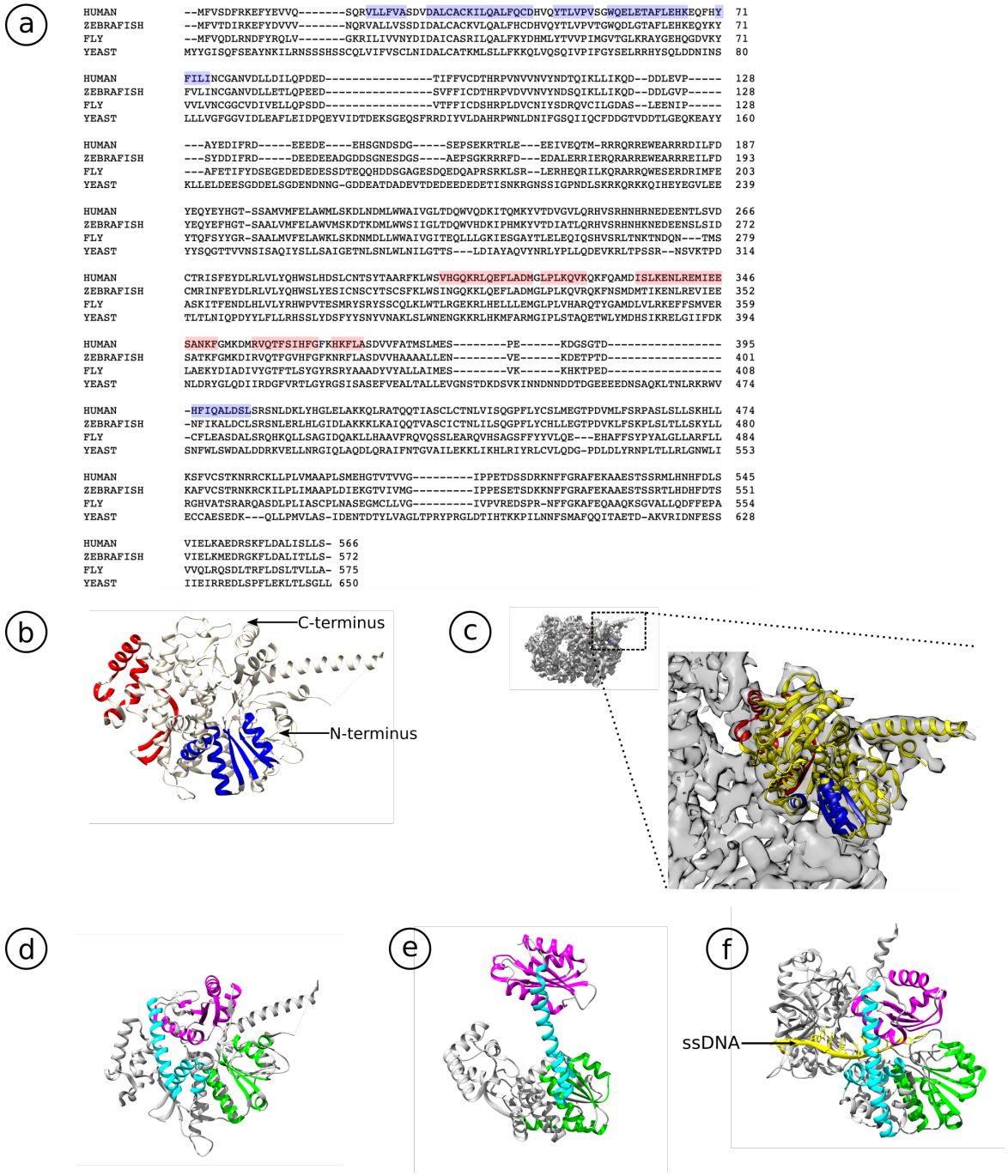



**Fig. 1.10 Human Cdc45 vs other species.** **(a)** Sequence alignment of eukaryotic Cdc45. The GINS-binding region highlighted in blue and Mcm2–7-binding in red. Collectively, they are termed CMG-Interacting Domain (CID). **(b)** Ribbon diagram of Cdc45 structure (PDB: 5DGO). Colour code as in (a). **(c)** Docking of human Cdc45's crystal structure into cryo-EM 3D map of *S. cerevisiae*'s CMG (EMDB ID: 3318). Inset: full 3D map of the CMG. Colour code as in (a), except the rest of Cdc45 is coloured in yellow to improve the contrast. **(d)** Structure of human Cdc45 highlighting the N-terminal DHH domain (green), C-terminal DHHA1 (magenta), and the connecting helical stem (cyan). **(e)** Structure of *T. kodakarensis* GAN (PDB: 5GHT). **(f)** Structure of *E. coli* RecJ bound to ssDNA (PDB: 5F56). Colour code in (e) and (f) the same as in (d). Adapted from (Pellegrini, 2017; Simon et al., 2016).

interactions between Cdc45's DHH–DHHA1 domains render any possible ssDNA binding site located at the interface inaccessible (figure 1.10d) (Pellegrini, 2017; Simon et al., 2016).

Given the importance of Cdc45 to ensure normal progression of cell cycle and replication, its recruitment to form the CMG helicase needs to be regulated. In accordance with the stepwise manner of the CMG helicase formation, Cdc45 is recruited first before the GINS. In order to carry out Cdc45 recruitment properly, a loading factor that is also a possible chaperone for Cdc45 is identified. Several studies in *S. cerevisiae* suggest this role is carried out by Sld3/Treslin (Masai et al., 2006; Sheu and Stillman, 2006; Tanaka et al., 2011a).

### 1.6.2 Sld3/Treslin in DNA replication

#### Sld3

In *S. cerevisiae*'s genetic screen for Dpb11's binding partner, Sld3 (Synthetic Lethality with Dpb11-3) was identified (Kamimura et al., 2001). The gene product of *SLD3* is essential for normal cellular growth and co-localises with replicating chromatin (Nakajima and Masukata, 2002). Sld3 was also found to interact with non-essential protein Sld7 (Tanaka et al., 2011a,b). Interestingly, the homologue of Sld7 could not be found in a related fungal species of *Schizosaccharomyces pombe* (fission yeast). Depletion of Sld7 caused slowing down of cell growth and delay in S-phase, but did not affect cell viability (Tanaka et al., 2011b).

Sld3 was confirmed to be important for DNA replication initiation, but not fork progression (Kanemaki and Labib, 2006; Yeeles et al., 2015). The possibility of Sld3 acting as a molecular chaperone for Cdc45 recruitment to the replication origin was supported by the observation that the complex of Sld3, Sld7, and Cdc45 (one of the CMG helicase components) is required for *S. cerevisiae*'s replication initiation (Masai et al., 2006; Sheu and Stillman, 2006; Tanaka et al., 2011a). Moreover, Sld3 was also reported to interact with Cdc45 at the DNA replication origin (Kamimura et al., 2001; Nakajima and Masukata, 2002; Tanaka et al., 2011a).

In a separate study to find the phosphorylation targets of S-CDK, Sld3 and Sld2 were identified to be the minimum S-CDK substrates for normal cell cycle progression (Tanaka et al., 2007; Zegerman and Diffley, 2007). When a *S. cerevisiae* strain expressed a Sld3–Dpb11 fusion and phosphomimetic Sld2, replication could still occur in the presence of CDK inhibitors (Zegerman and Diffley, 2007). Intriguingly, normal replication was also observed when an over-expressing allele of Cdc45 (so-called Jet1), phospho-mimetic Sld2, and Dpb11 in a S-CDK knock-out cell (Tanaka et al., 2007). The phosphorylation of Sld3 by S-CDK drives the formation of transient pre-loading complex (Sld3, Sld2, Dpb11, and

Pol$\varepsilon$) that subsequently completes CMG formation by recruiting the final component, GINS (Muramatsu et al., 2010).

S-CDK is not the only kinase that targets Sld3. Rad53 (Chk1 in human) is a major sensor kinase of the intra S-phase checkpoint system (section 1.4.6). Rad53 phosphorylation on Sld3 inhibits replication origin re-firing and re-replication during S-phase (Zegerman and Diffley, 2010). Rad53 was later shown to be important to stabilise replisome integrity upon fork stalling by targeting Sld3–Cdc45 complex (Can et al., 2019). Sld3 binding to Cdc45 is impaired after Rad53 phosphorylation, restricting the recruitment of Cdc45 to the DNA replication origin (Can et al., 2019; Yeeles et al., 2015; Zegerman and Diffley, 2010). This is one way how the cell cycle checkpoint ensures CMG helicase loading and activation occurs strictly during the G1 phase of cell cycle.

A crystal structure of the Sld3–Sld7 complex showed that the N-terminal domain of Sld3 forms a dimer with the N-terminal of Sld7 (Itou et al., 2015). Furthermore, the C-terminus of Sld7 was shown to form homodimer. This suggests a Cdc45 recruitment model of two copies of Sld7 binding to two copies of Sld3, which also acts as a chaperone for two copies of Cdc45 for depositing on the DNA-bound Mcm2–7 double hexamer. However, this Sld3-Sld7-Cdc45 complex has not been purified from *S. cerevisiae* or higher eukaryotes system. The binding of Cdc45 to Sld3 is further described in section 1.6.3.

It was also proposed that Sld3 level is low and hence becomes a limiting factor for replication initiation (Mantiero et al., 2011). Analysis of global protein expression in *S. cerevisiae* revealed that the number of Sld3 protein molecules per cell is roughly 250 to 300 (Ghaemmaghami et al., 2003; Tanaka et al., 2011a). Considering the number of potential replication origin sites in *S. cerevisiae* — approximately between 260–429 and the need of two molecules of Sld3 to fire one replication origin (Wyrick et al., 2001; Yabuki et al., 2002) — there is a need for regulated recycling of Sld3 for late replication origin firing. The regulation of Sld3/Treslin expression level and its network of interactions to ensure proper replication origin firing is not fully understood.

### Treslin is the functional orthologue of fungal Sld3

Finding the homologue of Sld3 in higher eukaryotes was not straightforward due to sequence divergence. Initially, the genetic screen effort focused on identifying a higher eukaryotic protein that shares Sld3's characteristics, such as: ability to bind to Dpb11/TopBP1 and Cdc45, while being a substrate for S-CDK. After a series of bioinformatic analyses and biochemical experiments, Treslin was confirmed as the higher eukaryote's homologue of Sld3 (Sanchez-Pulido et al., 2010).

Treslin (TopBP1-interacting Replication Stimulating Protein) was discovered in a co-immunoprecipitation experiment using *X. laevis*'s egg extract to identify TopBP1-interacting proteins (Kumagai et al., 2010). In a separate experiment using insertional mutation screens in *Danio rerio* (zebrafish), the same protein was identified and named as TICCR (TopBP1-Interacting Checkpoint and Replication Regulator) (Sansam et al., 2010). For the rest of the dissertation, this protein shall be referred to as Treslin.

Both studies confirmed Treslin's role during replication and its similarity with fungal Sld3. Treslin contains several predicted S-CDK sites, two of them — threonine 969 and serine 1001 — are conserved TopBP1-binding sites, similar to Sld3–Dpb11 interaction sites in *S. cerevisiae* (Boos et al., 2011, 2013; Kumagai et al., 2011). However, in Treslin there might be an additional and weak binding site for TopBP1 (Kumagai et al., 2011). The S1001 site was also shown to be a target for ATR kinase for intra-S phase checkpoint signalling (Hassan et al., 2013). Furthermore, a co-immunoprecipitation study in *X. laevis*'s egg extract coupled with mass spectrometry showed that Treslin forms a ternary complex with MTBP (Mdm2 Binding Protein) and TopBP1 (Boos et al., 2013). Throughout cell cycle, Treslin forms a constitutive complex with MTBP, via its MTBP-Binding Domain (MBD), similar to Sld3–Sld7 complex in *S. cerevisiae* (Boos et al., 2013; Tanaka et al., 2011a). MTBP was then proposed to be the functional orthologue of Sld7 protein. Similar to Sld7, MTBP is not essential for viability and not required for replication licensing (Boos et al., 2013). However, partial depletion of Treslin-MTBP-TopBP1 (mirroring Sld3-Sld7-Dpb11) slows down replication and lengthens S-phase of cell cycle (Boos et al., 2013).

As observed with Sld3, Treslin is also a target for the 'intra-S phase' checkpoint kinase Chk1 (Rad53 in *S. cerevisiae*) (Guo et al., 2015; Lopez-Mosqueda et al., 2010; Zegerman and Diffley, 2010). In Treslin, the Chk1 phosphorylation sites are located in a conserved domain called C-terminal region of Treslin (TRCT) encompassing residues 1810–1909 (Guo et al., 2015). The effect of Chk1 phosphorylation on Treslin is not fully understood, although it is possible that Chk1 negatively regulates Treslin-dependent recruitment of Cdc45 to block the CMG assembly when DNA damage is present during replication (Guo et al., 2015). Another regions in the C-terminal half, residue 1485–1660, was shown to interact with BRD2 and BRD4 proteins (Sansam et al., 2018). These two proteins bind to histone containing acetylated lysine residues to mediate interaction within the chromatin. Treslin–BRD4 formation is needed to ensure normal S-phase progression, highlighting various roles of Treslin during DNA replication (Sansam et al., 2018).

When Treslin is depleted in the cell, Cdc45 is no longer recruited to chromatin. This observation supports a mechanism in which the conserved role of Sld3/Treslin acting as Cdc45's loading factor or chaperone (Kumagai et al., 2010). Treslin down-regulation also

inhibits pre-initiation complex formation, possibly because Cdc45 recruitment to chromatin is disrupted (Kumagai et al., 2011). Apart from the replication process, Treslin is likely to play an additional role in the G2/M checkpoint (Sansam et al., 2010).

Despite their similar biological properties, confirming the sequence homology between Sld3 and Treslin was not trivial. Treslin is at least three times larger in size in comparison to Sld3 (figure 1.11a). Treslin possesses more domains, presumably to carry out more functions in higher eukaryotes. As an example, the N-terminal domain of Treslin is referred to as Conserved In Treslin (CIT) domain. This is conserved across higher eukaryotes, but not present in the fungal system (figure 1.11a). At last, a homology search using a limited sequence of Treslin (residue 717–792) finally confirmed sequence similarity with Sld3 (figure 1.11b) (Sanchez-Pulido et al., 2010). In a separate study, the S-CDK target sites of threonine/serine-proline (T/S-P) motifs were also shown to be conserved across eukaryotes (figure 1.11c) (Kumagai et al., 2011).

### 1.6.3 Sld3/Treslin and Cdc45 interaction

An essential part of Sld3/Treslin during DNA replication initiation is to act as a molecular chaperone or loading factor for Cdc45 deposition on to the DNA-bound Mcm2–7 (Kamimura et al., 2001; Tanaka et al., 2011a). It was demonstrated that the Sld3–Cdc45 interaction is required for DNA replication origin firing in *S. cerevisiae* (Masai et al., 2006; Tanaka et al., 2011a). However, neither the molecular mechanism of the association between Sld3/Treslin and Cdc45, nor the way the interaction promotes Cdc45 incorporation prior to CMG complex are understood.

Two-hybrid analysis in *S. cerevisiae* system showed that Sld3 interacts with Cdc45 throughout the cell cycle (Kamimura et al., 2001), a feature that was also observed in *S. pombe* (Nakajima and Masukata, 2002). The Sld3–Cdc45 interaction was also confirmed by recombinant co-expression of the two proteins (Itou et al., 2014). Unlike Cdc45, Sld3 is not required for replication fork progression (Kanemaki and Labib, 2006). However, the regulatory mechanism to disengage Sld3/Treslin after the completion of CMG assembly is currently unknown.

Rad53 phosphorylation inhibits the formation of the Sld3–Cdc45 complex which in turn prevents replication origin re-firing during S-phase of the cell cycle (Lopez-Mosqueda et al., 2010; Zegerman and Diffley, 2010). This could be bypassed by over-expressing Cdc45, which potentially removes the rate-limiting step of Sld3-dependent Cdc45 incorporation (Zegerman and Diffley, 2010). These observations can be treated as indirect proofs of Sld3's role as a chaperone or loading factor for Cdc45, which ensures proper replication origin

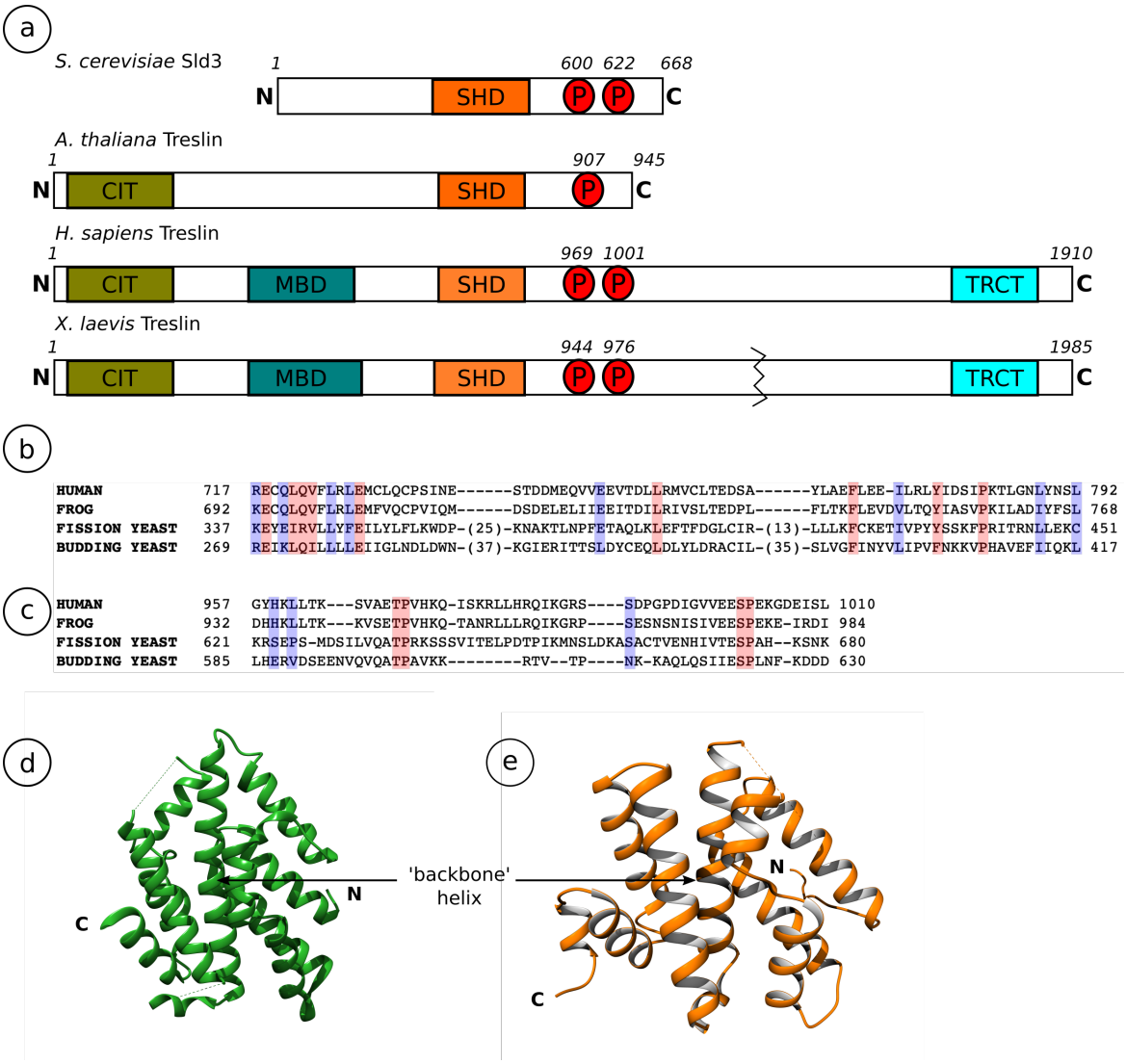



**Fig. 1.11 Sld3 and Treslin.** **(a)** Schematic of Sld3/Treslin across the species. CIT = Conserved in Treslin, MBD = MTBP Binding Domain, SHD = Sld3/Treslin Homology Domain, TRCT = Treslin C-Terminal, P = Conserved CDK phosphorylation sites (Itou et al., 2014). **(b)** Sequence alignment of Sld3/Treslin Homology Domain. Adapted from (Tanaka and Araki, 2013) **(c)** Sequence alignment of Sld3/Treslin conserved S-CDK phosphorylation sites. **(d)** Structure of *S. cerevisiae* (budding yeast) Sld3/Treslin Homology domain (SHD) (PDB: 3WI3, chain B). **(e)** Structure of human Treslin SHD (Simon *et al.*, unpublished).

firing by regulating CMG helicase assembly and activation. Once Sld3 is inhibited, CMG formation is also restricted unless Cdc45 is over-expressed.

Details of Cdc45–Treslin formation and regulation in higher eukaryotes are even less well-understood. Co-immuno-precipitation experiment using *X. laevis*'s egg extract demonstrated that the N-terminal half of the protein (residues 1–921) is sufficient to interact with over-expressed Cdc45 (Kumagai et al., 2011). The Cdc45-binding domain is likely to be the conserved domain called Sld3/Treslin Homology Domain (SHD) (figure 1.11a), a strong indication that the conserved role of Sld3/Treslin is to function as an interacting partner of Cdc45 (Itou et al., 2014; Kumagai et al., 2011). Nevertheless, the binding mechanism and regulation of Sld3/Treslin–Cdc45 complex is largely unclear.

There is no structural information on the Sld3/Treslin–Cdc45 complex, while there are some high-resolution structures of the individual proteins. The crystal structure of human Cdc45 was solved by the Pellegrini Lab, but it contained little information on how Cdc45 interacts with Sld3/Treslin (Simon et al., 2016). Meanwhile, the crystal structure of *S. cerevisiae* Sld3 SHD domain (residues 148–430) is available (figure 1.11d) (Itou et al., 2014). Sld3 adopts a compact, rhombus-shaped structure that comprises 12 helices. One central helix (annotated as the 7th $\alpha$-helix) acts as a 'backbone' of the domain, whereas other helices are placed around it through hydrophobic interactions. The structure highlighted that several residues that might interact with Cdc45, based on genetic analyses, were clustered on the same surface of the molecule. However, it is unclear whether the mutations in that region actually causes unfolding of the protein or are directly involved in Cdc45-binding. The crystal structure of human Treslin's SHD ($\text{Treslin}_{\text{SHD}}$) was also solved by the Pellegrini Lab, which adopts similar conformation as *S. cerevisiae*'s Sld3 (figure 1.11e) (Simon *et. al*, unpublished). The structure of Sld3/Treslin SHD alone offers little information on how the complex of Cdc45–Treslin is being formed. Thus, the structural analysis of Cdc45–Treslin complex is critical for understanding how complex formation occurs.

### 1.6.4 Sld3/Treslin and Mcm2–7 interaction

Together with Sld7 and Cdc45, Sld3 is recruited to chromatin in a DDK-dependent manner (Tanaka et al., 2011a). The DDK substrates during replication initiation are the Mcm2–7 subunits of Mcm2, Mcm4 and Mcm6 (Cho et al., 2006; Francis et al., 2009; Hardy et al., 1997; Sheu and Stillman, 2006, 2010). This phosphorylation event potentially creates a rigid domain which serves as Sld3 binding site (Abid Ali et al., 2017). Sld3 might also interact with Mcm2 in the absence of DDK, although the Sld3–Mcm2 interaction is greatly enhanced upon DDK phosphorylation (Herrera et al., 2015).

The Sld3–Mcm2–7 interaction is required for cell cycle progression and normal initiation of DNA replication (Deegan et al., 2016; Herrera et al., 2015). In *S. cerevisiae*, Sld3 is able to interact with Mcm2–7 independently of Cdc45 but not vice versa (Herrera et al., 2015). The Sld3–Mcm2–7 interaction is mediated by a region involving the SHD and the conserved TopBP1-binding sites (Deegan et al., 2016). Meanwhile, direct interaction of Mcm2–7–Treslin and how this possible interaction is being regulated in human system has not been reported. Isolation of Mcm2–7–Sld3/Treslin and Mcm2–7-Sld3/Treslin-Cdc45 complexes is difficult due to their transient interactions.

## 1.7 The objectives of the project

The general aim of this project was to contribute to our understanding on how CMG helicase is assembled at the DNA replication origin and how it becomes active. In particular, this work is focused on elucidating how Treslin participates in the process of CMG complex assembly.

As highlighted in the section 1.6.3 and 1.6.4, there are gaps in our understanding of Cdc45–Treslin and Mcm2–7–Treslin interactions in humans. Structural and regulatory information on these complexes are still mostly absent. In order to address these questions, a combination of biochemical, biophysical, and structural analyses were used. The methods and materials used for these experiments are described in chapter 2.

The characterisation of Cdc45–Treslin complex formation is discussed in chapter 3. Here, complex formation and regulation were studied by using purified recombinant human proteins. The interaction was confirmed by pulldown studies and semi-analytical size-exclusion chromatography. Furthermore, the binding affinity and the kinetics of the complex were also measured by Surface Plasmon Resonance (SPR) and Bio-Layer Interferometry (BLI). The structural analysis of this complex was also carried out by hydrogen-deuterium exchange mass spectrometry (HDX-MS) and cryo-electron microscopy (cryo-EM) analyses.

In chapter 4, the Mcm2–7–Treslin interaction is discussed. The direct interaction of Mcm2–7–Treslin is reported and a possible regulatory mechanism is also proposed. The studies involved pulldown, biophysical, and structural biology experiments. Treslin–TopBP1 interaction was also verified to be conserved across the eukaryotes. Furthermore, Treslin was found to contain a novel DNA-binding domain. Finally, chapter 5 summarises and discusses the implications of the results presented in this dissertation.

# Chapter 2

# Materials and methods

This chapter presents the details of reagents production and experimental protocols used to generate results presented in this dissertation. The sequence for all the oligonucleotides can be found in the appendix A.

## 2.1 Molecular Cloning

All the DNA plasmids described here encode human proteins, unless otherwise stated. PCR (Polymerase Chain Reaction) amplification was performed using a standard Phusion DNA polymerase (NEB) protocol, as summarised in the table below.

| **Stage** | **Description** | **Temp (°C)** | **Time (s)** |
|---|---|---|---|
| Stage 1 (1 ×) | Initial denaturation | 98 | 120 |
| Stage 2 (35 ×) | Denaturation | 98 | 20 |
| | Annealing | 55 - 60 | 20 |
| | Extension | 72 | 30 s / kb |
| Stage 3 (1 ×) | Additional extension | 72 | 300 |
| | Storage | 4 | ∞ |

Each of the resulting DNA plasmids was transformed into DH5$\alpha$ chemically competent *E. coli* cells (ThermoFisher) and plated on LB agar containing suitable antibiotics —kanamycin at 30 $\mu$g/mL for pRSFDuet1 vector or ampicillin at 100 $\mu$g/mL for other bacterial expression vectors described in this dissertation. For transformation of larger plasmid constructs (>8 kb in size), DNA plasmid was transformed into 10-beta chemically competent *E. coli* cells (NEB) and plated on LB agar containing gentamycin at 15 $\mu$g/mL, because the backbone of this large constructs contain gentamycin-resistant gene. All plated cells were incubated for 16–18 h at 37 °C.

Every clone was confirmed by DNA sequencing at the Sequencing Facility, Department of Biochemistry, University of Cambridge.

### 2.1.1 DNA plasmids for protein over-expression in bacteria

#### Cdc45 constructs

Both full-length and crystallisation (missing residue 154–164) constructs were cloned into a pRSFDuet1 bacterial expression vector (Novagen) via NdeI and XhoI restriction sites. The cDNA template was obtained from I.M.A.G.E clone 2964592 (Source BioSciences). These constructs were cloned by Dr Aline Simon and were already available in the lab before the start of the project (Simon et al., 2016).

These clones were subsequently used as templates to generate various Cdc45 alanine mutants using Quikchange PCR (Agilent) protocol with the following primers:

| | |
|---|---|
| Cdc45 K420A | *hCdc45-k420a-qcfw* and *hCdc45-k420a-qcrev* |
| Cdc45 K421A | *hCdc45-k421a-qcfw* and *hCdc45-k421a-qcrev* |
| Cdc45 K424A | *hCdc45-k424a-qcfw* and *hCdc45-k424a-qcrev* |

#### Treslin$_{SHD}$ constructs

All human Treslin constructs for bacterial cell expression were amplified from a synthetic DNA template encoding human Treslin residues 191–1102 codon optimised for *E. coli* over-expression (GeneArt, ThermoFisher).

Two constructs, comprising residues 626–813 and 503–934 (referred to as Treslin$_{SHD}$ or the Sld3/Treslin Homology Domain) were amplified using *hTreslin626-NcoI-fw* and *hTreslin813-XhoI-rev* (for 626–813) or *hTreslin503-NcoI-fw* and *hTreslin934-XhoI-rev* (for 503–934). The amplified sequences were cloned into pHAT4 and pGAT2 bacterial expression vectors via their NcoI and XhoI restriction sites. pHAT4 encodes a TEV-cleavable N-terminal hexahistidine tag, whereby pGAT2 encodes a Thrombin protease-cleavable N-terminal GST tag (Peranen et al., 1996).

To generate Treslin$_{SHD}$ with a TEV-cleavable N-terminal MBP tag and non-cleavable C-terminal octahistdine tag, the sequence 626–813 was amplified using *hTreslin626-EcoRI-fw* and *hTreslin813-XhoI-rev2* primers. The reverse primer contained sequence encoding a non-cleavable octahistidine tag. The amplified DNA was cloned into a modified pBAT4 bacterial expression vector containing a TEV-cleavable N-terminal MBP tag via unique EcoRI and XhoI sites. For vector map of the modified pBAT4 vector, see appendix A.

### Treslin$_{\text{SHD-ext}}$ constructs

Treslin$_{\text{SHD-ext}}$ is defined as the protein construct containing Treslin SHD domain + 'extension' region (located at the C-terminus end of SHD). In this dissertation, the main construct that represents the Treslin$_{\text{SHD-ext}}$ domain is the region covering 626–1053 residues. The sequence was amplified using the primers listed in the table below. The reverse primer contained sequence encoding a non-cleavable octahistidine tag. The amplified sequences were cloned into a modified pBAT4 vector containing a TEV-cleavable N-terminal MBP tag via unique EcoRI and XhoI sites.

| | |
|---|---|
| Treslin 626–881 | *hTreslin626-EcoRI-fw* and *hTreslin881-XhoI-rev* |
| Treslin 626–989 | *hTreslin626-EcoRI-fw* and *hTreslin989-XhoI-rev* |
| Treslin 626–1049 | *hTreslin626-EcoRI-fw* and *hTreslin1049-XhoI-rev* |
| Treslin 626–1053 | *hTreslin626-EcoRI-fw* and *hTreslin1053-XhoI-rev* |
| Treslin 626–1056 | *hTreslin626-EcoRI-fw* and *hTreslin1056-XhoI-rev* |
| Treslin 626–1059 | *hTreslin626-EcoRI-fw* and *hTreslin1059-XhoI-rev* |
| Treslin 626–1062 | *hTreslin626-EcoRI-fw* and *hTreslin1062-XhoI-rev* |
| Treslin 626–1065 | *hTreslin626-EcoRI-fw* and *hTreslin1065-XhoI-rev* |
| Treslin 626–1068 | *hTreslin626-EcoRI-fw* and *hTreslin1068-XhoI-rev* |
| Treslin 626–1071 | *hTreslin626-EcoRI-fw* and *hTreslin1071-XhoI-rev* |
| Treslin 626–1074 | *hTreslin626-EcoRI-fw* and *hTreslin1074-XhoI-rev* |
| Treslin 626–1077 | *hTreslin626-EcoRI-fw* and *hTreslin1077-XhoI-rev* |
| Treslin 626–1080 | *hTreslin626-EcoRI-fw* and *hTreslin1080-XhoI-rev* |
| Treslin 813–1053 | *hTreslin813-EcoRI-fw* and *hTreslin1053-XhoI-rev* |

### Treslin constructs for *in vivo* biotinylation

AviTag$^{\text{TM}}$ (Avidity LLC), amino acid sequence GLNDIFEAQKIEWHE, can be utilised for a specific lysine-biotinylation when engineered in a target protein. Using primer extension, AviTag$^{\text{TM}}$ was added to the N-termini of the Treslin$_{\text{SHD}}$ and Treslin$_{\text{SHD-ext}}$ constructs. The sequence was amplified using a common forward primer of *Avi-NcoI-fw* and reverse primer *hTreslin813-XhoI-rev2* (for residues 626–813) or *hTreslin1053-XhoI-rev* (for residues 626–1053). The templates for these DNA amplifications were the cloned Treslin 626–813 or 626–1053 in the modified pBAT4 vector containing TEV-cleavable N-terminal MBP tag, respectively. The amplified DNA was cloned into a pBAT4 bacterial expression vector via unique NcoI and XhoI sites.

### Treslin$_{\text{SHD-ext}}$ alanine mutants

Treslin$_{\text{SHD-ext}}$ clone (Treslin 626–1053, see subsection Treslin$_{\text{SHD-ext}}$ constructs) was used as a template to produce various single alanine mutants of Treslin$_{\text{SHD-ext}}$ using a QuikChange PCR (Agilent), Phusion polymerase (NEB), and primers listed in the table below.

| | |
|---|---|
| S820A | *hTreslin-s820a-qcfw* and *hTreslin-s820a-qcrev* |
| S838A | *hTreslin-s838a-qcfw* and *hTreslin-s838a-qcrev* |
| S923A | *hTreslin-s923a-qcfw* and *hTreslin-s923a-qcrev* |
| S969A | *hTreslin-s969a-qcfw* and *hTreslin-s969a-qcrev* |
| S1001A | *hTreslin-s1001a-qcfw* and *hTreslin-s1001a-qcrev* |
| S1013A | *hTreslin-s1013a-qcfw* and *hTreslin-s1013a-qcrev* |

To produce a construct of Treslin$_{\text{SHD-ext}}$ 6A mutants (S820A, S838A, S923A, S969A, S1001A, and S1013A), a synthetic DNA codon-optimised for *E. coli* expression encoding Treslin residues 626–1053 with 6A mutations was used as a template (GeneArt, ThermoFisher). The sequence was amplified using *hTreslin626-EcoRI-fw* and *hTreslin1053-XhoI-rev2* primers. The reverse primer contained a sequence encoding for a non-cleavable octahistidine tag. The amplified DNA was cloned into a modified pBAT4 vector containing a TEV-cleavable N-terminal MBP tag including unique EcoRI and XhoI sites.

### Treslin$_{\text{SHD-ext}}$ phosphoserine incorporation

A tRNA/tRNA synthase system developed at Jason Chin Laboratory allows incorporation of phosphoserine to the translated protein when recognising the 'amber' (TAG) codon (Rogerson et al., 2015). This method was used to incorporate phosphoserine in two sites (T969 and S1001) of Treslin by using cloned Treslin$_{\text{SHD-ext}}$ as a template. The construct was generated by a QuikChange PCR protocol (Agilent) using primers listed below.

| | |
|---|---|
| T969pS | *hTreslin-t969ps-qcfw* and *hTreslin-t969ps-qcrev* |
| S100pS | *hTreslin-s1001ps-qcfw* and *hTreslin-s1001ps-qcrev* |

### Treslin$_{\text{SHD-ext}}$ phosphomimic mutant

Treslin$_{\text{SHD-ext}}$ clone (Treslin 626–1053, see subsection Treslin$_{\text{SHD-ext}}$ constructs) was used as a template to produce the phosphomimic mutant of Treslin in amino acid residues T969 and S1001 by QuikChange PCR (Agilent) and primers listed below.

| | |
|---|---|
| T969E | *hTreslin-t969e-qcfw* and *hTreslin-t969e-qcrev* |
| S1001D | *hTreslin-s1001d-qcfw* and *hTreslin-s1001d-qcrev* |

### S-CDK constructs

The constructs of human S-phase Cyclin-Dependent Kinase (S-CDK), comprising Cyclin A and CDK2, were gifts from Dr Philip Zegerman (Gurdon Institute, Cambridge).

Cyclin A with C-terminal hexahistidine tag was cloned into a pETDuet1 vector via NcoI and BlpI sites. Meanwhile, CDK2 was cloned into pGEX vector via BamHI sites to enable N-terminal tagging with a GST tag.

### DDK constructs

The constructs for human Dbf4-Dependent Cdc7 Kinase (DDK) were gifts from Dr Peter Cherepanov (Francis Crick Institute, London).

The constructs for Cdc7 were either full-length or containing N-terminus deletion (missing residues 1–36), and cloned into a pRSFDuet1 vector via NcoI and XhoI sites. Meanwhile for Dbf4 constructs —either encoding full-length protein or residues 210–350 — were cloned into a pCDFDuet1 vector with sequence encoding for a 3C protease-cleavable hexahistdine tag via NcoI and XhoI restriction sites (Hughes et al., 2012).

### TopBP1 NTD construct

TopBP1 NTD (Topoisomerase II Binding Protein 1, N-terminal Domain) containing BRCT0 to BRCT2 repeats, corresponding to residues 1–290, was cloned into a pHAT4 bacterial expression vector via NcoI and SalI restriction sites. pHAT4 encodes an N-terminal 3C protease-cleavable hexahistdine tag. Primers *hTopBP1-NcoI-fw* and *hTopBP1-290-SalI-rev* were used to amplify the cDNA template obtained from I.M.A.G.E clone 8991925 (Source BioSciences).

### *E. coli* BirA ligase construct

BirA is an *E. coli* ligase that biotinylates specific lysine residues in a biotin-acceptor motif. For biophysical measurements described in this dissertation, N-terminal AviTag$^{TM}$ (an optimised biotin-acceptor motif) was engineered to the N-terminus of Treslin to allow the biotinylation reaction.

The full-length *E. coli* BirA ligase cloned into a pMAT11 bacterial expression vector was cloned by Dr Joseph Maman and already available in the lab before the start of the project. This construct was used as a template to generate two constructs: MBP-tagged or Strep-MBP-tagged BirA using the primers listed below. Both were cloned into a pRSFDuet1 bacterial expression vector (Novagen) via EcoRI and SalI restriction sites.

| | |
|---|---|
| MBP-BirA | *MBP-BirA-EcoRI-fw* and *BirA-SalI-rev* |
| Strep-MBP-BirA | *Strep-MBP-BirA-EcoRI-fw* and *BirA-SaI-rev* |

### 2.1.2 DNA plasmids for protein over-expression in mammalian and insect cells

#### Mcm2–7 constructs for mammalian cell expression

DNA constructs encoding all subunits of human MCM, Mcm2–7, with TEV-cleavable C-terminal octahistidine tag (on Mcm4) and 3×FLAG tag (either on Mcm2 or Mcm5) were cloned by Dr Neil Rzechorzek and already present in the lab before the start of the project (Rzechorzek et al., 2020).

*MCM2–7* genes were synthesised as DNA fragments (GeneArt, ThermoFisher) and codon optimised for over-expression in HEK-293F mammalian cells. Each of the gene string was subsequently used as a template for amplification by PCR reaction and cloned into a pACEMam1 mammalian expression vector (Geneva Biotech) via BamHI and HindIII sites.

To combine two or more genes together into a single pACEMam1 vector, the 'recipient' vector was digested by BstXI whereas the 'donor' was cut by ICeuI (Berger and Craig, 2011). The DNA products were then ligated. The end results were a single vector containing *MCM2*, *MCM3*, and *MCM5* protein-encoding genes and a second vector that has *MCM4*, *MCM6*, and *MCM7*.

DNA sequence encoding a TwinStrep tag flanked with either NgoMIV or SalI sites was amplified from *StrepII-string* using primers listed below and used to replace the 3×FLAG tag on either *MCM2* or *MCM5*. For *MCM2*, the digestion was performed using NgoMIV whereby for *MCM5* with SalI.

| | |
|---|---|
| TwinStrep on Mcm2 | *StrepII-Fw-NgoMIV* and *StrepII-Rev-NgoMIV* |
| TwinStrep on Mcm5 | *StrepII-Fw-SalI* and *StrepII-Fw-SalI* |

#### Treslin constructs for mammalian and insect cell expression

A construct of full-length human Treslin with C-terminal 3×FLAG and S-tags cloned into a pCDNA5/TO mammalian expression vector was a kind gift from Professor William Dunphy (California Institute of Technology, California, US) (Kumagai et al., 2010). This clone was used as a template to perform two-step PCR reactions (using primers listed below), to generate full-length Treslin or Treslin residues 1–1053 ($\text{Treslin}_{1-1053}$) constructs with TEV-cleavable C-terminal TwinStrep tag. The resulting DNA products were cloned into a pACEMam1 mammalian expression vector (Geneva Biotech) via unique BamHI and BssHII restriction sites.

Cloned full-length human Treslin and $\text{Treslin}_{1-1053}$ constructs were re-cloned into a pACEBac1 insect cell expression vector (Geneva Biotech) via BamHI and EcoRI restriction

| | |
|---|---|
| TreslinFL (first PCR) | *hTreslin-BamHI-fw* and *hTreslin-1st-rev* |
| TreslinFL (second PCR) | *hTreslin-BamHi-fw* and *hTreslin-BssHII-rev* |
| $\text{Treslin}_{1-1053}$ (first PCR) | *hTreslin-BamHI-fw* and *hTreslin1053-1st-rev* |
| $\text{Treslin}_{1-1053}$ (second PCR) | *hTreslin-BamHI-fw* and *hTreslin1053-BssHII-rev* |

sites. Both pACEMam1 and pACEBac1 vectors contain identical multiple cloning sites (MCS).

### MTBP constructs for mammalian cell expression

The full-length human MTBP construct was cloned into a pACEMam1 mammalian expression vector (Geneva Biotech) via BamHI and XbaI restriction sites. The cDNA template was obtained from I.M.A.G.E clone 4476886 (Source BioSciences). The primers for amplification were *hMTBP-BamHI-fw* and *hMTBP-XbaI-rev*. The reverse primer sequence also encoded a TEV-cleavable octahistidine tag.

The resulting construct was re-cloned into a pACEBac1 insect cell expression vector (Geneva Biotech) via the same restriction sites.

### Bacmid and virus generation for Treslin and MTBP insect cell expression

Each aliquot of DH10-Bac$^{TM}$ chemically competent *E. coli* cell (ThermoFisher) was transformed with 10 ng of either pACEBac1 Treslin or MTBP construct. The cells were recovered in 450 $\mu$L SOC media for 5 h at 37 °C. 50 $\mu$L of the cells were plated onto LB agar plate containing 50 $\mu$g/mL kanamycin, 15 $\mu$g/mL gentamycin, 10 $\mu$g/mL tetracycline, 200 $\mu$g/mL X-gal, and 40 $\mu$g/mL IPTG. The plates were incubated for 48 h at 37 °C until the blue and white colonies could be distinguished. To ensure clonal purity, one white colony was picked and re-streaked on a fresh LB agar containing the same compositions as above. The plate was incubated for additional 48 h at 37 °C.

Each one of the white colonies from the re-streaked plate was picked to inoculate 5 mL of 2×YT supplemented with all the antibiotics listed in the earlier paragraph. After the cells were harvested, alkaline lysis was performed to isolate the DNA. This was followed by isopropanol DNA precipitation, washing with 70% (v/v) ethanol, and air-drying the DNA pellet. The resulting DNA pellet was resuspended in 40 $\mu$L TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA). The identity of the recombinant bacmid was confirmed by standard PCR reaction using *M13fw* and *M13rev* primers. The successful recombinant bacmid yielded two bands on 1% (w/v) agarose gel electrophoresis: a band corresponding to the gene-of-interest and 2300 bp band corresponding to the bacmid backbone.

The transfection reagent to generate the P1 virus was prepared by mixing 100 $\mu$L of Insect-XPRESS (Lonza) media and 1 $\mu$g of recombinant bacmid. At the same time, solution

of 100 $\mu$L of Insect-XPRESS (Lonza) media + 8 $\mu$L of Cellfectin II was also prepared and incubated for 1 h at room temperature. Afterwards, both solutions were mixed. 210 $\mu$L of this mixture was added drop-wise to 1–3 mL of Sf9 insect cells (cell count = $8 \times 10^5$) inside a 6-well plate for transfection. The plate was incubated for 5 h at 27 °C. After incubation, all the liquid from the plate was aspirated and replaced by 2 mL of Insect-XPRESS media supplemented with 10% (v/v) heat-inactivated FBS. The plate was incubated for 5 days at 27 °C. The cell culture was harvested by centrifuging at 500 *g* for 5 min and the resulting supernatant was transferred to a fresh tube. This was P1 virus.

The virus was amplified further by adding 0.5 mL of P1 virus to 25 mL of Sf9 insect cells at $2 \times 10^6$ cells/mL. The cells were left shaking at 120 rpm at 27 °C. The cell culture density was maintained within 2–2.5 $\times 10^6$ cells/mL by adding the Insect-XPRESS media to ensure exponential growth phase throughout the days. When the cells stopped dividing and the viability dropped below 80% (assessed by Trypan Blue staining), the cells were harvested. The supernatant was recovered, mixed with 5% heat-inactivated FBS, and transferred to fresh tube. This was P2 virus.

$2 \times 25$ mL of Sf9 insect cells at $2 \times 10^6$ cells/mL were prepared. One flask was treated with 100 $\mu$L of P2 virus whereby the other was not. Both flasks were left shaking at 120 rpm for 24 h at 27 °C. Afterwards, the cell density for each flask was calculated. From these results, the fraction of cells inhibited was calculated using the following equation

$$x = 1 - \frac{A - 2 \times 10^6}{B - 2 \times 10^6} \tag{2.1}$$

where $x$ = fraction of inhibited cells, A = density of infected cells, and B = density of untreated cells.

The Multiplicity of Infection (MOI) was then calculated using this polynomial approximation to the Poisson function using this following equation

$$MOI = 5.0725x^4 - 5.758x^3 + 2.6479x^2 + 0.7425x \tag{2.2}$$

where $x$ = fraction of cells inhibited. From this actual MOI, the actual concentration of the viral stock was derived from this following equation.

$$V = 1 \times 10^9 \times \frac{MOI}{2} \tag{2.3}$$

where $V$ = concentration of viral stock and MOI = multiplicity of infection.

## 2.2 Protein over-expression in bacteria

The general protocols for bacterial transformation, cell culture, and cell harvest are largely similar for all proteins. Any difference in protein over-expression is specified in the upcoming subsections.

BL21 Rosetta2 DE3 chemically competent *E. coli* cells (Novagen) were transformed with 50 ng of each DNA construct. In case of Treslin *in vivo* biotinylation, equimolar amounts of a plasmid encoding for BirA were also co-transformed. The cells were left for 30 min on ice, heat-shocked for 45 s at 42 °C, left for 2 min on ice, and recovered by shaking at 215 rpm for 1 h at 37 °C with 450 $\mu$L of SOC media (Formedium). The cells were plated on LB agar containing chloramphenicol at 34 $\mu$g/mL and additional antibiotic(s): kanamycin at 30 $\mu$g/mL for pRSFDuet1, spectinomycin at 100 $\mu$g/mL for pCDFDuet1, and ampicillin at 100 $\mu$g/mL for other bacterial expression vectors. The plated cells were incubated for 16–18 h at 37 °C.

**Table 2.1** Type and amount of media

| **Protein expressed** | **Type of media** | **Amount of media** |
|---|---|---|
| Cdc45 and its mutants | Turbo broth | 2 × 1 L |
| Treslin SHD and biotinylation | 2 × YT | 2 × 1 L |
| Treslin SHD-ext, its mutants, and biotinylation | | 12 × 1 L |
| Cyclin A | | 2 × 1 L |
| CDK2 | | 2 × 1 L |
| Cdc7 and Dbf4 (DDK) | LB | 2 × 1 L |
| TopBP1 NTD | 2 × YT | 4 ×1 L |

The resulting colonies were used to inoculate the media inside a two-litre baffled flask supplemented with the same antibiotics as in the transformation step (for the type and total amount of media, see table 2.1). The cultures were left shaking at 200 rpm at 37 °C. When the optical density of the culture at 600 nm ($OD_{600}$) reached 0.9, final concentration of 0.5 mM IPTG was added to each litre of culture (except for S-CDK constructs, 0.1 mM IPTG was used). In the case of *in vivo* biotinylation of the AviTag recombinant proteins, D-biotin (Sigma) at final concentration of 25 $\mu$M dissolved in pre-warmed 1×PBS was also added. The cultures were left shaking at 200 rpm for additional 16 h at 20 °C.

The cells were harvested by centrifuging at 4,000 *g* for 10 min at 25 °C (Beckman Avanti J-20XP). The pellet from a 2-L culture was resuspended in 25 mL of resuspension buffer specified at table 2.2. The resuspended cells were sonicated on ice with 5 s ON / 10 s OFF at 60% amplitude for total time of 3 min (Vibra-Cell SONICS) followed by centrifugation at 39,500 *g* for 1 h at 4 °C. The clarified lysate was subsequently filtered using 5 and 0.45 $\mu$m syringe filters.

**Table 2.2** Resuspension buffer

| Protein | Resuspension buffer |
|---|---|
| Cdc45 and its mutants | 20 mM HEPES pH 8.0, 160 mM NaCl, 1 mM DTT |
| $Treslin_{SHD}$ | 20 mM HEPES pH 8.0, 160 mM NaCl, 20 mM imidazole |
| Biotinylated $Treslin_{SHD}$ | 20 mM HEPES pH 8.0, 160 mM NaCl, 20 mM imidazole |
| $Treslin_{SHD\text{-}ext}$ and its mutants | 20 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole |
| Biotinylated $Treslin_{SHD\text{-}ext}$ | 20 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole |
| Cyclin A and CDK2 (S-CDK) | 20 mM HEPES pH 7.5, 300 mM NaCl |
| Cdc7 and Dbf4 (DDK) | 50 mM $Na_2HPO_4$ pH 7.5, 300 mM NaCl, 2 mM NaF 0.025% (v/v) IGEPAL, 10% (w/v) glycerol, 20 mM imidazole |
| TopBP1 NTD | 20 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole |

Notes: All buffers were supplemented with one tablet of EDTA-Free protease inhibitor (Sigma) per 100 mL. The pH of the 1M HEPES stock was adjusted with NaOH.

### 2.2.1 Cdc45 purification

The clarified and filtered lysate was applied to a 5 mL HiTrap Q HP anion exchange column (GE Healthcare) equilibrated in *Cdc45 Buffer-A* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT). The column was washed with 25 column volumes (CV) of *Cdc45 Buffer-A*, before eluting the bound proteins over a 25 CV gradient from 0-100% of *Cdc45 Buffer-B* (20 mM HEPES-NaOH pH 8.0, 1 M NaCl, 1 mM DTT). The pooled eluate was diluted to a final concentration of 160 mM NaCl.

The pooled eluate was applied to a 5 mL HiTrap Heparin HP column (GE Healthcare). The buffers, column washing, elution, and salt concentration adjustment steps were done as the HiTrap Q HP anion exchange step.

The pooled fractions from the HiTrap Heparin HP step were applied to a 6 mL ResourceQ anion exchange column (GE Healthcare) equilibrated in *Cdc45 Buffer-A*. The column was washed with 25 CV of *Cdc45 Buffer-A* before eluting the bound proteins over 30 CV gradient from 0–100% of *Cdc45 Buffer-B*.

The peak fraction from the ResourceQ anion exchange step was applied to a Superdex75 16/600 size-exclusion column (GE Healthcare) equilibrated in *Cdc45 Buffer-A*. The peak fractions were pooled, concentrated to ~15 mg/mL, snap frozen in liquid nitrogen, and stored at -80 °C.

### 2.2.2 His-GST tagged $Treslin_{SHD}$ purification

For summary of buffer compositions used in all Treslin purifications, refer to table 2.3.

The lysate was loaded onto a gravity column containing 2 mL Ni-NTA agarose resin (Qiagen) equilibrated in *Treslin Buffer-A* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl,

**Table 2.3** Buffers used for Treslin purification

| Treslin Buffer | Buffer compositions |
|---|---|
| Buffer A | 20 mM HEPES pH 8.0, 160 mM NaCl, 20 mM imidazole |
| Buffer B | 20 mM HEPES pH 8.0, 160 mM NaCl, 1 mM DTT |
| Buffer C | 20 mM HEPES pH 8.0, 1 M NaCl, 1 mM DTT |
| Buffer D | 20 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole |
| Buffer E | 20 mM HEPES pH 8.0, 160 mM NaCl, 1 mM DTT, 5 %(w/v) glycerol |
| Buffer F | 20 mM HEPES pH 8.0, 1 M NaCl, 1 mM DTT, 5 %(w/v) glycerol |
| Buffer G | 20 mM HEPES pH 8.0, 300 mM NaCl, 20 mM imidazole, 5%(w/v) glycerol |
| Buffer H | 50 mM HEPES pH 8.0, 300 mM NaCl, 1 mM DTT, 5%(w/v) glycerol |
| Buffer J | 50 mM Tris-Cl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 5%(w/v) glycerol |
| Buffer K | 20 mM HEPES pH 8.0, 300 mM NaCl, 1 mM DTT, 5%(w/v) glycerol |
| Buffer L | 50 mM HEPES pH 8.0, 300 mM NaCl, 5%(w/v) glycerol, 0.1%(v/v) IGEPAL |
| Buffer M | 50 mM HEPES pH 8.0, 300 mM NaCl, 5%(w/v) glycerol |

Notes: All buffers with DTT were prepared fresh on the day of the experiment. The pH of the 1 M HEPES stock was adjusted with NaOH. The pH of the potassium glutamate (K-glutamate) was adjusted with KOH.

20 mM imidazole). The column was washed with 20 CV of *Treslin Buffer-A* before eluting the bound proteins with 5 CV of *Treslin Buffer-A* + 300 mM imidazole.

The eluate from the Ni-NTA affinity step was pooled and applied to a gravity column containing 2 mL glutathione agarose resin (Cube Biotech) equilibrated with *Treslin Buffer-B* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT). The column was capped and left on a roller mixer for 1 h at 4 °C. The column was emptied by gravity, washed with 20 CV of *Treslin Buffer-B*, and the bound proteins were eluted with 5 CV *Treslin Buffer-B* + 20 mM reduced glutathione, pH adjusted to 8.0 with NaOH dissolved in *Treslin Buffer-B*.

The eluate from the glutathione affinity step was pooled, concentrated to ~2 mL, and applied on a Superdex200 16/600 size-exclusion column (GE healthcare) equilibrated in *Treslin Buffer-B*. The peak fractions were pooled, concentrated to ~2 mg/mL (for $Treslin_{503–934}$) or ~8 mg/mL (for $Treslin_{626–813}$), snap frozen in liquid nitrogen, and stored at -80 °C.

### 2.2.3 $Treslin_{SHD}$ purification

The clarified and filtered lysate was loaded onto a gravity column containing 2 mL Ni-NTA agarose resin (Qiagen) equilibrated in *Treslin Buffer-A* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 20 mM imidazole). The column was washed with 20 CV of *Treslin Buffer-A* followed by eluting the bound proteins with 5 CV of *Treslin Buffer-A* + 300 mM imidazole. The eluate was pooled, mixed with 1:50 (v/v) His-tagged TEV protease (at 2 mg/mL), and dialysed overnight at 4 °C (SnakeSkin 10K MWCO, ThermoFisher) against *Treslin Buffer-B* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT).

The sample was passed through a gravity column containing 1-mL Ni-NTA agarose resin to capture the cleaved His-tags and His-tagged TEV protease. The flow through from the column was pooled, concentrated to 2 mL, and applied to a Superdex75 16/600 column (GE Healthcare) equilibrated in *Treslin Buffer-B*. The peak fractions were pooled, concentrated to ~40 mg/mL, flash frozen in liquid nitrogen, and kept at -80 °C.

### 2.2.4 Biotinylated Treslin$_{SHD}$ purification

The initial purification steps were as described in section 2.2.3, except that the TEV protease addition and the dialysis steps were omitted. The eluate was applied to a 5 mL HiTrap Q HP anion exchange column (GE Healthcare) equilibrated in *Treslin Buffer-B*. The column was washed with 25 CV of *Treslin Buffer-B*, before eluting the bound proteins over a 25 CV gradient from 0–100% of *Treslin Buffer-C*.

The eluate from the anion-exchange step was concentrated to 2 mL and applied to a Superdex75 16/600 size-exclusion column (GE Healthcare) equilibrated in *Treslin Buffer-B*. The peak fractions from the size-exclusion step were pooled, concentrated to ~23 mg/mL, flash frozen in liquid nitrogen, and stored at -80 °C.

### 2.2.5 Treslin$_{SHD\text{-}ext}$ protein expression trial

Various constructs of Treslin$_{SHD\text{-}ext}$ were used for transformation. For small scale trial expression, 250 mL of culture was used. The induction, cell culture, harvest, and lysate clarification steps were as described on section 2.2, adjusting the volume of resuspension buffer proportionally.

The total and soluble fractions were then separated by SDS-PAGE gel and subjected to Western blot using anti-His (section 2.16). The construct expressing Treslin residues 626–1053 was better expressed than the rest and subsequently used for larger scale expression and purification.

### 2.2.6 Treslin$_{SHD\text{-}ext}$ phosphoserine incorporation protein expression trial

Constructs of Treslin$_{SHD\text{-}ext}$ with T969pS, S1001pS, or both were used for transformation as described in section 2.2, except using Rosetta2 DE3 strain of chemically competent *E. coli* cells that expressed tRNA/tRNA synthase system to incorporate phosphoserine while reading the 'amber' codon (Rogerson et al., 2015). A 1 L culture was used for each construct. All the cell culture, harvest, and lysate clarification steps were performed as described in the

beginning of section 2.2. During the induction step, a final concentration of 1 mM IPTG and 0.2 mM phosphoserine (Sigma) were added to each litre of culture.

Upon cell lysis and initial Ni-NTA affinity purification, the resulting protein was very poorly expressed. The purification was subsequently abandoned.

### 2.2.7 Treslin$_{\text{SHD-ext}}$ and its mutants purification

The lysate was loaded onto a gravity column containing 6 mL Ni-NTA agarose resin (Qiagen) equilibrated in *Buffer Treslin-D* (20 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 20 mM imidazole). The column washed with 20 CV of *Treslin Buffer-D* before eluting the bound proteins with 5 CV of *Treslin Buffer-D* + 300 mM imidazole. The eluates were pooled and adjusted to 160 mM NaCl concentration.

The pooled eluate from the Ni-NTA capture step was applied to a 5 mL HiTrap Heparin HP column (GE Healthcare) equilibrated in *Treslin Buffer-B* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT). The column was washed with 25 CV of *Treslin Buffer-B*, and followed by eluting the bound proteins over 25 CV gradient from 0–100% of *Treslin Buffer-C* (20 mM HEPES-NaOH pH 8.0, 1 M NaCl, 1 mM DTT). To remove the N-terminal MBP tag, 1:50 (v/v) of His-tagged TEV protease at 2 mg/mL was added to the pooled eluates and left incubating on a roller mixer for 15 h at 4 °C.

The sample was filtered using 0.22 $\mu$m syringe filter before loading it onto a gravity column containing 1 mL Ni-NTA agarose resin (Qiagen) equilibrated in *Treslin Buffer-A*. The column was washed with 20 CV of *Treslin Buffer-A* followed by elution of bound proteins using *Treslin Buffer-A* + 300 mM imidazole.

The eluates were pooled, concentrated to ~2 mL, and applied to a Superdex75 16/600 size-exclusion column (GE Healthcare) equilibrated in *Treslin Buffer-J* (50 mM Tris-Cl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 5% (w/v) glycerol). The peak fractions were pooled, concentrated to ~5 mg/mL, snap frozen in liquid nitrogen, and stored at -80 °C.

When purifying the Treslin$_{\text{SHD-ext}}$ with the MBP tag intact, the TEV protease addition and Ni-NTA rebinding were skipped. The pooled eluates from the Heparin capture step were applied to a Superdex200 16/600 size-exclusion column (GE Healthcare) equilibrated in *Treslin Buffer-K* (20 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 1 mM DTT, 5% (w/v) glycerol). The peak fractions were pooled, concentrated to ~10 mg/mL, snap frozen in liquid nitrogen, and kept at -80 °C.

### 2.2.8 Biotinylated Treslin$_{\text{SHD-ext}}$ purification

All the steps were the same as described in section 2.2.7, except the TEV protease addition and Ni-NTA rebinding steps were omitted.

### 2.2.9 S-CDK purification

Separate lysates of CDK2 and CyclinA were mixed and left incubating at roller mixer for 1 h at 4 °C. 2.5 mL glutathione agarose resin (Cube Biotech) equilibrated in *CDK Buffer-A* (20 mM HEPES-NaOH pH 7.5, 300 mM NaCl) was added and left incubating for additional hour at 4 °C. The mixture was then applied onto a gravity column, washed with 20 CV of *CDK Buffer-A*, before eluting the bound proteins with 5 CV of *CDK Buffer-A* + 20 mM reduced glutathione, pH adjusted to 7.5 with NaOH dissolved in *CDK Buffer-A*.

The eluate was loaded onto a gravity column containing 1 mL of Ni-NTA agarose resin (Qiagen) equilibrated in *CDK Buffer-B* (20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, 20 mM imidazole). The column was washed with 20 CV of *CDK Buffer-B* before eluting the bound proteins with 5 CV of *CDK Buffer-B* + 300 mM imidazole.

The eluted fractions were pooled and subjected to a PD10 buffer exchange column (GE Healthcare) equilibrated in *CDK Buffer-A*. The fractions were snap frozen in liquid nitrogen and stored at -80 °C.

### 2.2.10 DDK purification

The purification was described by Hughes *et al* with some modifications (Hughes et al., 2012). The clarified and filtered lysate was applied onto a gravity column containing 3 mL Ni-NTA agarose resin (Qiagen) equilibrated in *DDK Buffer-A* (50 mM $NaH_2PO_4$ pH 7.5, 300 mM NaCl, 0.025%(v/v) IGEPAL, 10% (w/v) glycerol, 20 mM imidazole). The column was washed with 20 CV of *DDK Buffer-A* before eluting the bound proteins with 10 CV of *DDK Buffer-A* + 300 mM imidazole.

The eluate from the Ni-NTA capture step was pooled and applied to a 5 mL HiTrap Q HP anion exchange column (GE Healthcare) equilibrated in *DDK Buffer-B* (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 10%(w/v) glycerol, 1 mM DTT). The column was washed with 25 CV of *DDK Buffer-B* and then followed by eluting the target proteins over 25 CV gradient from 0–100% of *DDK Buffer-C* (same as *DDK Buffer-B*, except 1 M NaCl).

The peak fractions from the anion exchange step were pooled, concentrated to final volume of 2 mL, and applied to a Superdex200 16/600 size-exclusion chromatography column (GE Healthcare) equilibrated in *DDK Buffer-D* (50 mM Tris-HCl pH 7.5, 150 mM

NaCl, 2 mM NaF, 10%(w/v) glycerol). The peak fractions from the gel filtration were pooled, concentrated to ~2 mg/mL, flash frozen in liquid nitrogen, and stored at -80 °C.

### 2.2.11 TopBP1 NTD purification

The filtered lysate was applied to a gravity column containing 3 mL Ni-NTA agarose resin (Qiagen) equilibrated in *TopBP1 Buffer-A* (20 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 20 mM imidazole). The column was washed with 25 CV *TopBP1 Buffer-A*, followed by addition of 5 CV *TopBP1 Buffer-A* + 300 mM imidazole to elute the bound proteins.

The eluate from the Ni-NTA step was pooled and concentrated to 2 mL before applying to a Superdex75 16/600 size-exclusion column (GE Healthcare) equilibrated in *TopBP1 Buffer-B* (20 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 1 mM DTT). Peak fractions from the size-exclusion step were pooled, concentrated to ~20 mg/mL, snap frozen in liquid nitrogen, and kept at -80 °C.

## 2.3 Protein over-expression in mammalian cells

### 2.3.1 Mcm2–7 purification

3× 400 mL of FreeStyle™ HEK-293F resuspension cells (ThermoFisher) were grown in the FreeStyle media (Gibco, Invitrogen) to a cell density of $1 \times 10^6$ cells/mL before transfection. The transfecting reaction was prepared by mixing together FreeStyle media (10% volume of the total culture), 0.6 $\mu$g of each DNA construct per million of cells, and 2× weight (with respect to the total DNA) of polyethylenimine (PEI) (Sigma). The transfection mixture was mixed for 10 s by vortexing, incubated for 15 min at room temperature, and poured directly into the flask. 3 h post transfection, a final concentration of filtered 4 mM valproic acid (Sigma) dissolved in deionised water was added to the cell culture. The culture was incubated for 4 days at 125 rpm, 70% humidity, 6% $CO_2$ level, and 37 °C.

The cells were harvested by spinning at 500 *g* (Beckman Avanti J-20XP) for 5 min at 25 °C. Each cell pellet was washed with 30 mL of 1× PBS and re-spun at 500 *g* (Eppendorf 5430R) for 5 min at 25 °C. The cell pellets were resuspended in 25 mL of *MCM Buffer-A* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 20 mM imidazole, 5% (w/v) glycerol) supplemented with 1 tablet/100 mL of EDTA-free protease inhibitor (Sigma) and 2 units of benzonase (Sigma, E8263). The resuspended cells were sonicated on ice with 5 s ON / 10 s OFF, 55 % amplitude, total time 2 min (Vibra-Cell SONICS). The lysate was clarified at 41,000 *g* for 1 h at 4 °C.

The clarified lysate was filtered consecutively using 5, 0.45, and 0.22 $\mu$m syringe filters before applying it to a gravity column containing 3 mL Ni-NTA agarose resin (Qiagen) equilibrated in *MCM Buffer-A*. The column was washed with 20 CV of *MCM Buffer-A* before eluting the bound proteins with 5 CV of *MCM Buffer-A* + 300 mM imidazole.

The pooled eluate from the Ni-NTA affinity step was applied, at flow rate of 0.2 mL/min, to a 1 mL StrepTrap HP column (GE Healthcare) equilibrated in *MCM Buffer-B* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 5%(w/v) glycerol). The column was washed with 20 CV of *MCM Buffer-B* before eluting the bound proteins in reverse flow direction with 5 CV of *MCM Buffer-B* + 15 mM d-desthiobiotin (Sigma), with pH adjusted to 7.5 by KOH dissolved in *MCM Buffer-B*. The peak fractions were flash frozen in liquid nitrogen, and stored at -80 °C.

### 2.3.2 Treslin and Treslin–MTBP complex over-expression trial

The cell culture, transfection, and harvest steps are similar to those described in subsection 2.3.1. The resuspension buffer used was *Treslin Buffer-L* (50 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 5% (w/v) glycerol, 0.1% (v/v) IGEPAL).

#### Treslin purification attempt

The clarified lysate was filtered using 5, 0.45, and 0.22 $\mu$m syringe filters. The filtered lysate was applied to a gravity column containing 2 mL StrepTactin agarose resin (IBA), equilibrated in *Treslin Buffer-M* (50 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 5%(w/v) glycerol). The column was capped and left incubating in roller mixer for 1 h at 4 °C. The column was emptied by gravity and washed with 20 CV *Treslin Buffer-M*. The bound proteins were eluted in 5 CV *Treslin Buffer-M* + 15 mM d-desthiobiotin, with pH adjusted to 8.0 by NaOH dissolved in *Treslin Buffer-M*.

#### Treslin–MTBP complex purification attempt

The clarified lysate was filtered using 5, 0.45, and 0.22 $\mu$m syringe filters. The filtered lysate was applied to a gravity column containing 3 mL Ni-NTA agarose resin (Qiagen), equilibrated in *Treslin Buffer-M* with 20 mM imidazole. The column was washed with 20 CV *Treslin Buffer-M* + 20 mM imidazole, before eluting the bound proteins with 5 CV of *Treslin Buffer-M* + 300 mM imidazole.

The pooled eluate from the Ni-NTA capture step was applied to a gravity column containing 1 mL StrepTactin agarose resin (IBA), equilibrated in *Treslin Buffer-M*. The column was capped and left incubating in roller mixer for 1 h at 4 °C. The column was

emptied by gravity and washed with 20 CV of *Treslin Buffer-M*. The proteins were eluted in 5 CV of *Treslin Buffer-M* + 15 mM d-desthiobiotin, with pH adjusted to 8.0 by NaOH dissolved in *Treslin Buffer-M*.

## 2.4 Protein over-expression in insect cells

### 2.4.1 Treslin and Treslin–MTBP purification attempts

2 × 1 L of Sf9 insect cells (Invitrogen) were grown in Insect-XPRESS media (Lonza) to a cell density of 2 × $10^6$ cells/mL before transfection. The transfection was done by adding each of the the P2 virus carrying recombinant bacmid of Treslin or MTBP so the Multiplicity of Infection (MOI) is 2. The cells were incubated for 3 days at 120 rpm and 27 °C.

The cells were harvested by spinning at 500 *g* (Beckman Avanti J-20XP) for 5 min at 25 °C. The cell pellet was resuspended with 30 mL 1 × PBS and re-spun at 500 *g* (Eppendorf 5430R). The cell pellet was resuspended in *Treslin Buffer-M* (50 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 5% (w/v) glycerol, 0.1% (v/v) IGEPAL) supplemented with 1 tablet/100 mL of EDTA-free protease inhibitor (Sigma) and 2 units of benzonase (Sigma, E8263). The resuspended cells were lysed by sonication on ice with 5 s ON / 10 s OFF, 55% amplitude, total time 2 min (Vibra-Cell SONICS). The lysate was clarified by centrifuging at 41,000 *g* for 1 h at 4 °C.

The purification steps for Treslin and Treslin–MTBP were exactly the same as described in the subsection 2.3.2.

## 2.5 Pulldown experiments

### 2.5.1 GST pulldown of Treslin and Cdc45

10 $\mu$M of purified His-GST tagged Treslin recombinant proteins or His-GST protein (as a control) were bound to 60 $\mu$L glutathione agarose beads (Cube Biotech) in a 1.5 mL Eppendorf tube equilibrated in *GST Pulldown Buffer-A* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT, 0.1 mg/mL BSA). The tube was left at roller mixer for 1 h at 4 °C. The beads were then washed with 3 × 1 mL of buffer, pelleting the beads in between each wash by spinning at 2,000 *g* for 1 min at 4 °C (Eppendorf).

Purified Cdc45 at final concentration of 20 $\mu$M was added to the beads and the tubes were left at roller mixer for 1 h at 4 °C. Afterwards, the beads were washed as described before, except using *GST Pulldown Buffer-B* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM

DTT). The bound protein complexes were eluted by adding 200 $\mu$L of *Pulldown Buffer-B* + 20 mM reduced glutathione (Sigma), pH 7.5 adjusted by NaOH dissolved in *GST Pulldown Buffer B* and left on the roller mixer for 10 min at 4 °C. The supernatant was recovered after pelleting the beads by centrifuging at 2,000 *g* for 2 min at 4 °C and analysed by SDS-PAGE.

### 2.5.2 GST pulldown of Treslin and Cdc45 in various salt concentrations

The steps were the same as in section 2.5.1. The difference is that after binding of the His-GST tagged Treslin$_{SHD}$, the beads were washed with varying concentrations of NaCl (160, 200, 300, and 400 mM NaCl) in the buffer.

### 2.5.3 MBP pulldown of Treslin and Cdc45

10 $\mu$M of purified MBP-tagged Treslin (and its various mutants) proteins or MBP protein (as a control) were bound to 60 $\mu$L amylose agarose beads (NEB) in a 1.5 mL Eppendorf tube equilibrated in *MBP Pulldown Buffer-A* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT) + 0.1 mg/mL BSA. The tube was left on a roller mixer for 1 h at 4 °C. The beads were then washed with 3 × 1 mL of buffer, pelleting the beads in between each wash by spinning at 2,000 *g* for 1 min at 4 °C (Eppendorf).

Purified S-CDK at final concentration of 250 nM + 1 mM ATP + 0.1 mg/mL BSA were then added to the beads (or without any S-CDK, as a control for non-phosphorylated sample) and incubated on a roller mixer for 1 h at room temperature. Afterwards, the beads were washed as described previously.

20 $\mu$M of purified Cdc45 + 0.1 mg/mL BSA was mixed with the beads and incubated on a roller mixer for 1 h at 4 °C. Following the washing steps as described earlier, the bound proteins were eluted by adding 200 $\mu$L of *MBP Pulldown Buffer-A* + 20 mM d-maltose (Sigma) and incubated on a roller mixer for 10 min at 4 °C. The supernatant was recovered after pelleting the beads by centrifuging at 2,000 *g* for 2 min at 4 °C and analysed by SDS-PAGE.

### 2.5.4 MBP pulldown of Treslin and Mcm2–7

In the case of Mcm2–7 dephosphorylation prior to the pulldown, purified Mcm2–7 was treated with 1.2 $\mu$g of $\lambda$-phosphatase (SantaCruz, sc-200312) using *$\lambda$-PP buffer* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 5 % (w/v) glycerol, 0.5 mM $MnCl_2$) for 30 min at 30 °C.

5 $\mu$M of purified recombinant MBP-tagged Treslin or MBP (as a control) were bound to 40 $\mu$L of amylose agarose beads (NEB) in a 1.5 mL Eppendorf tube, equilibrated in *MCM Pulldown Buffer-A* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 5% (w/v) glycerol) + 0.1 mg/mL BSA. The tube was left at roller mixer for 1 h at at 4 °C. The beads were then washed 3 × 1 mL of buffer, pelleting the beads in between each wash by spinning at 2,000 *g* for 1 min at 4 °C.

Purified S-CDK at final concentration of 250 nM + 1 mM ATP + 0.1 mg/mL BSA were added to the beads (as a control for non-phosphorylated sample, S-CDK was omitted). The tubes were left on a roller mixer for 1 h at room temperature. Subsequently, the beads were washed as described previously.

Purified Mcm2–7 at final concentration of 0.4 $\mu$M + 0.1 mg/mL BSA was added to the beads. The tubes were left on a roller mixer for 1 h at at 4 °C. The washing steps were repeated as described previously. The bound proteins were eluted by adding 125 $\mu$L of *MCM Pulldown Buffer-A* + 20 mM d-maltose and left on a roller mixer for 10 min at 4 °C. The supernatant was recovered after pelleting the beads by centrifuging at 2,000 *g* for 2 min at 4 °C, and analysed by SDS-PAGE.

In the case case of DDK treatment of Mcm2–7, a final concentration of purified DDK at 100 nM + 1 mM ATP + 0.1 mg/mL BSA was added to the Mcm2–7 before addition to the beads (as a negative control, DDK was omitted). The tubes were incubated for 45 min at 30 °C.

### 2.5.5 StrepTactin pulldown of Mcm2–7 and Treslin

The Mcm2–7 dephosphorylation step, if required, was the same as described in the section 2.5.4.

0.5 $\mu$M of purified Mcm2–7 with a TwinStrep tag on Mcm2 and an octahistidine on Mcm4 or Strep-tagged BirA (as a control) were bound to 30 $\mu$L of StrepTactin agarose resin (IBA) in a 1.5 mL Eppendorf tube, equilibrated in *MCM Pulldown Buffer-A* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 5%(w/v) glycerol) + 0.1 mg/mL BSA. The tube was left on a roller mixer for 1 h at 4 °C. The beads were then washed 2 × 1 mL of buffer, pelleting the beads in between each wash by spinning at 2,000 *g* for 1 min at 4 °C .

While the Mcm2–7 was incubating with the beads, purified untagged $Treslin_{SHD\text{-}ext}$ at 5 $\mu$M was phosphorylated by adding 250 nM of purified S-CDK + 1 mM ATP. As a negative control, another reaction without S-CDK was also prepared. Both reactions were left for 30 min at room temperature. The phosphorylated (or non-phosphorylated) $Treslin_{SHD\text{-}ext}$ was added to the beads containing the Mcm2–7 with 0.1 mg/mL BSA. The tubes were left on a roller mixer for 1 h at 4 °C.

After washing the beads as previously described, bound proteins were eluted by adding 125 $\mu$L of *MCM Pulldown Buffer-B* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$, 15 mM d-desthiobiotin). The pH of this elution buffer was adjusted to 7.5 by adding KOH dissolved in *MCM Pulldown Buffer-B*. The supernatant was recovered after pelleting the beads at 2,000 *g* for 2 min at 4 °C, and analysed by SDS-PAGE.

### 2.5.6 StrepTactin pulldown of the Mcm2–7 and Treslin–Cdc45 complex

All the steps were as described in section 2.5.5, except 5 $\mu$M of purified Cdc45 was added to the reaction buffer following the $Treslin_{SHD\text{-}ext}$ phosphorylation step. The complex was pre-formed by mixing equimolar amount of purified Cdc45 and $Treslin_{SHD\text{-}ext}$.

### 2.5.7 MBP pulldown of Treslin and TopBP1 NTD

All the steps were as described in section 2.5.3, except that 20 $\mu$M of purified TopBP1 NTD was added following the S-CDK phosphorylation step.

## 2.6 Size-exclusion chromatography - multi-angle light scattering (SEC-MALS)

The Size-Exclusion Chromatography coupled with Multi-Angle Light Scattering (SEC-MALS) technique allows the determination of the absolute molecular weight (Mw) of protein species. The SEC part separates particles based on their hydrodynamic volume, with larger molecules eluted earlier than smaller ones. This process depends mainly on the porosity of the gel inside the column. The molecular weight determination by SEC alone is inherently inaccurate as it depends on protein standards and assumption of perfect globularity of the proteins and absence of interactions between the proteins and column resin. When SEC is being coupled with MALS, the SEC merely separates the protein species. The basic principle of MALS is that light scatters into different paths when passing through matter. To analyse the molecular weight of a molecule, the scattering of laser light is detected through detectors located at multiple different angles. Across a protein peak, light scattering detected at multiple angles is combined with protein concentration information (from in-line differential refractometer) and analysed by a Zimm plot to determine the absolute molecular weight. The main benefit of SEC-MALS is that it can analyse samples containing multiple protein species and achieve mass determination independent of shape (Some et al., 2019; Wyatt, 1993).

For the SEC-MALS experiments described in this dissertation, the column outlet was fed through in-line UV detector, eight Dawn Heleos II multi-angle light scattering detectors (Wyatt Technology), and finally Optilab T-rEX differential refractometer (Wyatt Technology). The data were collected and analysed by the Astra6 software (Wyatt Technology), with molecular masses calculated across eluted protein peaks through extrapolation from Zimm plots using a refractive index increment (dn/dc) value of 0.1850 mL/g; quoted molecular masses relate to the overall mass calculation across a peak (Harding and Jumel, 1998; Wyatt, 1993). Before each experimental run, BSA at 2 mg/mL (ThermoFisher) was used to calibrate the system. The predicted Mw for each protein sample was calculated by submitting the amino acid sequence into the ExPASy ProtParam tool at https://web.expasy.org/protparam/ (last accessed at 26th of August 2019).

### 2.6.1 SEC-MALS analysis of the Cdc45–Treslin complex

A Superdex200 Increase 10/300 size-exclusion chromatography column (GE Healthcare) was equilibrated overnight with *MALS Buffer-A* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT) at flow rate 0.5 mL/min. 70 $\mu$L sample of purified Cdc45 at 130 $\mu$M or Treslin$_{\text{SHD}}$ at 650 $\mu$M or Cdc45-Treslin$_{\text{SHD}}$ complex (by pre-incubating 130 $\mu$M Cdc45 with 650 $\mu$M Treslin$_{\text{SHD}}$ for 30 min on ice) was applied to the column. Before applying, each sample was centrifuged at 14,000 *g* for 5 min at 4 °C.

For experiments involving the Cdc45–Treslin$_{\text{SHD-ext}}$ complex, the concentration used for Cdc45 was 50 $\mu$M and Treslin$_{\text{SHD-ext}}$ at 50 $\mu$M. However, the buffer used to analyse Treslin$_{\text{SHD-ext}}$ by itself was *MALS Buffer-B* (20 mM HEPES-NaOH pH 8.0, 300 mM NaCl, 1 mM DTT). For the complex, the buffer used was *MALS Buffer-A*.

### 2.6.2 SEC-MALS analysis of the Mcm2–7 complex

A Superose6 Increase 3.2/300 size-exclusion column (GE Healthcare) was equilibrated overnight with *MALS Buffer-C* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$) at flow rate of 0.04 mL/min. 60 $\mu$L of purified Mcm2–7 at 1 $\mu$M with TwinStrep tag either on Mcm2 or Mcm5 was centrifuged at 14,000 *g* for 5 min at 4 °C before loading onto the column.

## 2.7 Semi-analytical size-exclusion chromatography

### 2.7.1 Cdc45–Treslin complex

The buffer compositions and amount of the proteins used were as described in section 2.6.1 without coupling the gel filtration system to the MALS detectors.

### 2.7.2 Mcm2–7 complex

The steps were as described in section 2.6.2, except the MALS step was omitted.

### 2.7.3 Treslin–TopBP1 NTD complex

50 $\mu$M of MBP-tagged $\text{Treslin}_{\text{SHD}}$ was incubated with 0.5 $\mu$M of S-CDK for 30 min at 25 °C. Purified TopBP1 NTD at 100 $\mu$M was added to the sample and left incubated for 30 min on ice. 125 $\mu$L of the sample was applied to a Superdex20 10/300 size-exclusion chromatography column (GE Healthcare) equilibrated with *TopBP1 Buffer-C* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT). The sample was centrifuged at 14,000 *g* for 5 min at 4 °C before application to the column.

## 2.8 Bio-Layer Interferometry (BLI)

Bio-Layer Interferometry (BLI) is an optical technique that measures the interference patterns between light waves. The light is reflected by two layers at the tip of the biosensors, a biocompatible matrix where the immobilised ligand is attached and an internal reference layer. The reflected light beams can interfere positively or negatively with each other at different wavelength and this is detected by a charge-coupled device (CCD) array detector. When the ligand is immobilised in the biosensor, it forms a two-dimensional layer on the surface. As this biosensor is introduced into the analyte, the thickness at the tip of the layer changes. This resulted in change of effective distance between two layers and hence shift in interference pattern of reflected light. This change of the pattern is changed as a function of the optical thickness in the tip of the biosensor. The spectral shift is then monitored as a change in wavelength in nanometer (figure 2.1). This read-out can be monitored in real time and used to deduce kinetic data. In addition, the equilibrium binding assay can provide the dissociation constant ($K_d$) of the complex (Kumaraswamy and Tobias, 2015; Sultana and Lee, 2015).

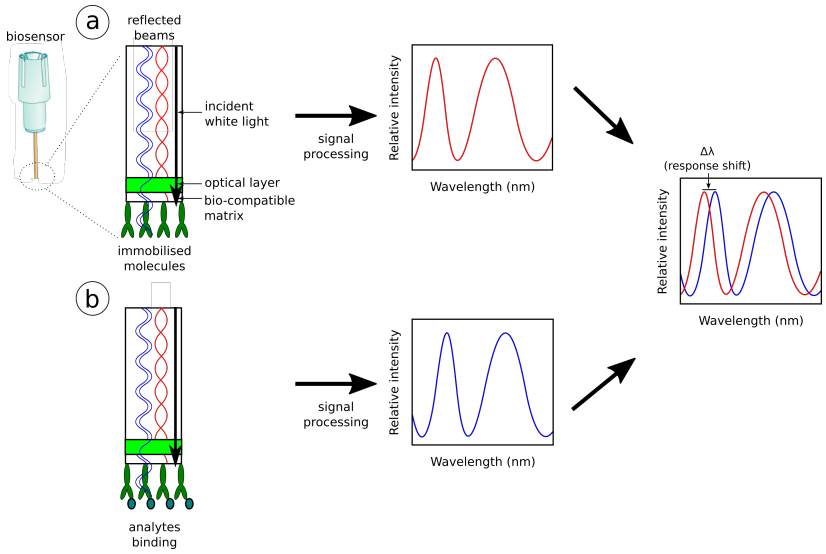



**Fig. 2.1 Bio-Layer Interferometry (BLI) detection principle.** Adapted from OCTET Red96 Pall ForteBio handbook. **(a)** Schematic of data processing when the white incident light is reflected by the biocompatible matrix and optical layers. The immobilised ligands are represented as green-coloured antibody. **(b)** Same as (a), except when the analytes are bound to the immobilised ligand. **(c)** The interference pattern change is monitored as response shift (nm). See section 2.8 for more details.

Bio-Layer Interferometry experiments were performed in the Octet Red96 system (Pall ForteBio). The reactions were shaking at 1000 rpm at 30 °C throughout the experiments. *BLI Buffer* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 0.1% (v/v) Tween 20, 0.1 mg/mL BSA) was used throughout for dilution, sensor washing, baseline recording, association, and dissociation steps. 200 $\mu$L of buffer was used in each well of the black 96-well plate (Greiner).

For analysis of Cdc45 and Treslin$_{\text{SHD}}$, anti-Strep biosensors (Pall ForteBio) were saturated with 10 nM of biotinylated Treslin$_{\text{SHD}}$ for 60 s and washed for 60 s in buffer. After 60 s immersion in the wells containing buffer to get the baseline recording, the loaded sensors were immersed for 60 s into the wells containing purified Cdc45 at concentrations of 30 $\mu$M to 0.469 $\mu$M. To perform dissociation step, the biosensors were then immersed for 60 s into the same wells used for baseline recording. At the end of each assay, the biosensors were regenerated by using 10 mM glycine pH 2.0 and immediately washed with the *BLI Buffer*. The results were analysed using the Pall ForteBio DataAnalysis software.

For studies of Cdc45 and biotinylated Treslin$_{\text{SHD-ext}}$, an anti-Strep biosensors (Pall ForteBio) were saturated with 15 nM biotinylated Treslin$_{\text{SHD-ext}}$ for 60 s. All the subsequents steps were the same as above, except both the association and dissociation steps were carried out for 240 s. Unfortunately, the Cdc45 (analyte) did not dissociate from the sensor, even when prolonging the dissociation step to 1 h.

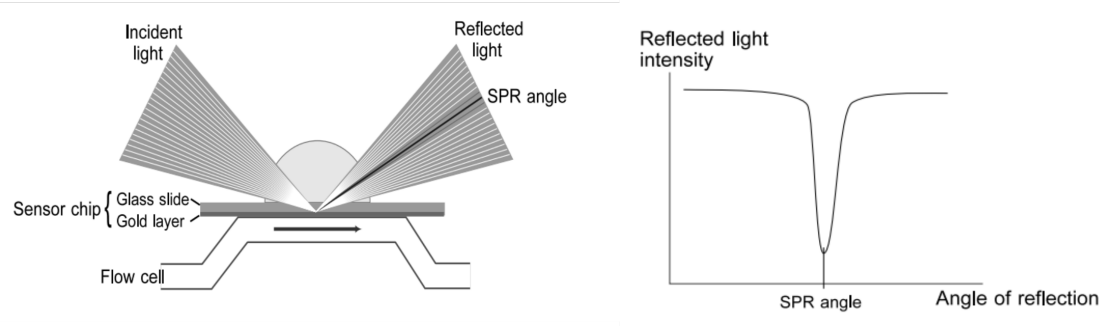



**Fig. 2.2 Surface Plasmon Resonance (SPR) detection principle** Adapted from Biacore Handbook. See section 2.9 for more details.

## 2.9 Surface Plasmon Resonance (SPR)

The basis of detecting protein–protein interactions using this technique depends on an optical phenomenon called Surface Plasmon Resonance (SPR). SPR is affected by the refractive indexes of the two media in between the conducting non-magnetic metal surface. SPR occurs when the light incident on the metal couples to the oscillation of the free electrons (called plasmons) at the metal surface. This creates an electromagnetic charge field, called evanescent wave, that penetrates through the metal layer and into the medium with a lower refractive index (usually the non-illuminated side of the metal). The evanescent wave exponentially decays with the distance from the metal layer (Mol and Fischer, 2010; Van der Merwe, 2001).

SPR takes place in total internal reflection (TIR) condition, where the intensity of reflected light equals to the incident light. During plasmon resonance, the intensity of reflected light decreases sharply at the defined angle of incidence called the SPR angle (figure 2.2). The SPR angle is dependent on various factors, such as wavelength, optical properties of the prism, and conducting metal. When these parameters are kept constant, the change in SPR angle shift is only caused by the changes in refractive index of the media around the metal layer. If monitored over time, the increase in proteins bound at the surface causes subsequent increase in the SPR angle, which is detected as a change in the position of light intensity minimum by the diode array. The measurement output is usually given in Resonance Units (RU), in which 1000 RU corresponds to a 0.1° shift in SPR angle (Mol and Fischer, 2010; Van der Merwe, 2001).

The SPR experiments were performed using Biacore T200 (GE Healthcare) with flow rate of 30 $\mu$L/min and at temperature of 30 °C (analysis) and 10 °C (sample chamber). *SPR buffer* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT, 0.005% (v/v) Tween 20) was used throughout the experiments.

500 $\mu$L of 100 nM biotinylated MBP-tagged Treslin was used as an immobilised ligand and prepared in the same batch of *SPR buffer*. This solution was used to couple the biotinylated ligand on the two-dimensional, gold layered SPR sensor chip SAP (XanTec Bioanalytics) by using an automated pulse-mode injections at 30 $\mu$L/min into each flow cell to achieve targeted resonance unit (RU) for each ligand substrate (400 RU for $\text{Treslin}_{\text{SHD}}$, 650 RU for $\text{Treslin}_{\text{SHD-ext}}$). Prior to protein immobilisation, the sensor chip was pre-conditioned with *Conditioning buffer* (50 mM NaOH, 1 M NaCl) for 10 min, as recommended by the sensor chip manufacturer.

For the experimental run, various dilutions of Cdc45 (from 50 to 0.390625 $\mu$M, including negative control without Cdc45) in the same batch of *SPR buffer* were injected in random order. Each run was started with *SPR buffer* injection for 60 s (baseline), followed by protein injection for 60 s (association), and then back to *SPR buffer* for 60 s (dissociation). Regeneration step was not performed as the signal went back to baseline level during dissociation step, indicating complete dissociation of the analytes. All the data points in sensorgrams were exported and plotted in ProFit software (Quantumsoft).

To obtain the steady state affinity graph, the flat section of the sensorgram at the end of injection step was averaged and plotted against concentration of analyte. The data points were fitted with the steady state affinity model (equation 2.4) below using Levenberg-Marquardt algorithm in ProFit software (Quantumsoft).

$$y = \frac{(1/K_{\text{d}}) \cdot x \cdot R_{\text{max}}}{1 + (1/K_{\text{d}}) \cdot x \cdot n} \tag{2.4}$$

where $y$ = the equilibrium binding, $x$ = the molar concentration of analyte, $K_d$ = the dissociation equilibrium constant (M), $R_{max}$ = theoretical maximum binding capacity, and $n$ = steric interference factor when experimental $\text{R}_{\text{max}}$ deviates from theoretical $\text{R}_{\text{max}}$.

To obtain the dissociation rate, the observed dissociation in sensorgram was fitted using the following dissociation first order decay model,

$$y = R \cdot (1 - e^{-k \cdot (x - Ti)}) + P \tag{2.5}$$

where $y$ = the observed resonance (RU), $x$ = time (s), $k$ = kinetic rate constant ($\text{s}^{-1}$), $R$ = $\Delta$RU for dissociation phase (RU), $P$ = plateau after dissociation step (RU), and $Ti$ = injection time (s).

Dissociation rate from all concentration of the analyte was averaged. The value was then used to derive the association rate using the following equation 2.6

$$K_{\text{on}} = \frac{K_{\text{off}}}{K_{\text{d}}} \tag{2.6}$$

where $K_{on}$ = association rate ($M^{-1}$ $s^{-1}$), $K_{off}$ = dissociation rate ($s^{-1}$), and $K_d$ = dissociation equilibrium constant (M).

## 2.10 Electrophoretic Mobility Shift Assay (EMSA)

Initial DNA binding studies of $Treslin_{SHD\text{-}ext}$ and $Treslin_{ext}$ were performed using an Electrophoretic Mobility Shift Assay (EMSA). When DNA forms a complex with a protein, the resulting complex compared to the free DNA migrates slower in the agarose gel. The binding constant ($K_d$) can be estimated by the concentration in which half of the amount of DNA is shifted in the gel.

The DNA and protein samples were dissolved in *EMSA Buffer* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl) and kept in black 1 5-mL tubes (Eppendorf). A 6FAM-labelled DNA was used at a final concentration of 10 nM. The DNA sequences used in these experiments can be found in the appendix A. The DNA was then mixed with various concentration of protein (total volume 20 $\mu$L) to achieve final concentrations of 35 $\mu$M to 170 nM. As a control, no protein was added to the free DNA.

The samples were run in a 1% agarose gel at 0.5×TBE buffer at 40 V for 90 min at 4 °C. Before applying the samples, the gel was pre-run for 30 min.

## 2.11 Fluorescence anisotropy

Fluorescence anisotropy is a phenomenon that occurs when plane polarised light excites a fluorophore, resulting in light emission intensities that vary between different axes of polarisation (a term called depolarisation).

In solution, the free fluorophores are tumbling randomly following Brownian motion. In this case, the rotational diffusion is faster than the rate of light emission that results in a fully depolarised light and anisotropy equals to zero. When the rate of tumbling is slower than the rate of emission, the light becomes polarised. This phenomenon can be used to obtain the binding affinity constant ($K_d$) of protein to DNA, provided that the size of protein–DNA complex is significantly larger than the free fluorophore. Normally, the DNA is being labelled by a fluorophore. Upon protein binding, the resulting complex tumbles more slowly than the free labelled DNA species. The polarisation of the emitted light causes an increase in anisotropy, which can be viewed as a result of protein–DNA complex formation (figure 2.3).

Fluorescence anisotropy is defined as:

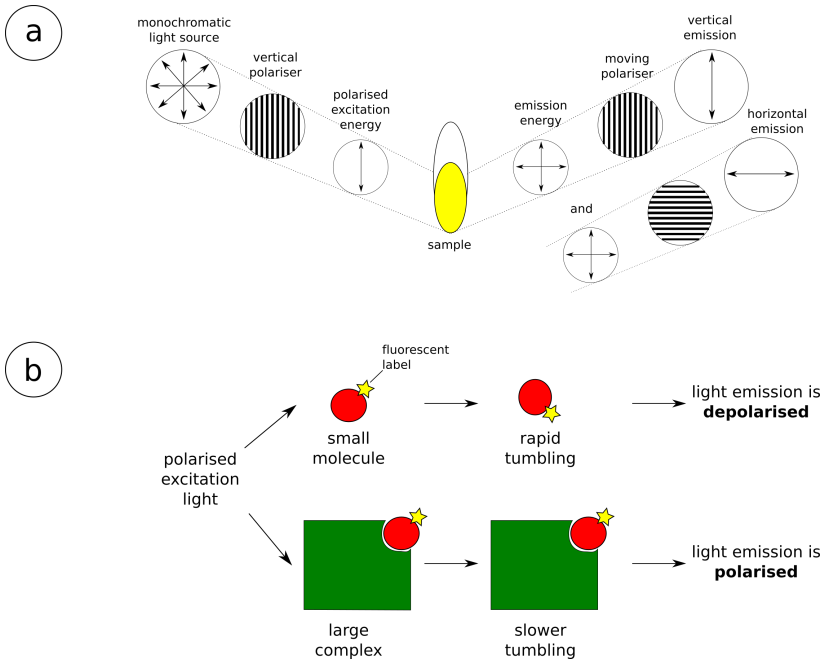



**Fig. 2.3 Fluorescence anisotropy principle.** Adapted from Invitrogen Technical Resource Guide. **(a)** Representation of fluorescence anisotropy detection. **(b)** Anisotropy difference between small molecule and larger complex. See section 2.11 for more details.

$$A = \frac{I_{\mathrm{v}} - I_{\mathrm{h}}}{I_{\mathrm{v}} + 2I_{\mathrm{h}}} \tag{2.7}$$

where $I_{\mathrm{v}}$ and $I_{\mathrm{h}}$ are vertically and horizontally polarised emission intensities, respectively and A is the anisotropy, independent of the total intensity of the emission light.

The DNA binding studies of $\mathrm{Treslin}_{\mathrm{SHD\text{-}ext}}$ was performed by using fluorescence anisotropy. 5' fluorescent (6-FAM)-labelled DNA was used at a final concentration of 20 nM. The DNA sequences used in the experiments can be found in the appendix A. Various concentrations of 40 $\mu$L purified protein (done by serial dilution from 40 $\mu$M to 78 nM) was used for the experiments. All the protein and DNA samples were dissolved in *FP Buffer* (20 mM HEPES-NaOH pH 8.0, 160 mM NaCl) and subsequently kept in black 1.5 mL Eppendorf tubes.

All anisotropy measurements, using an excitation wavelength of 485 nm and an emission wavelength of 520 nm at 25 °C, were performed in PHERAstar FS plate reader (BMG Labtech). Prior to the experiment, the gain and focal weight of the instrument was adjusted by using 20 nM fluorescein (target reading 35 mP). During the measurement, the samples were kept in black 384-well plate (Greiner).

The data point was fitted with anisotropy ligand depletion model (equation 2.8) using Robust algorithm in ProFit software (Quantumsoft) to obtain the dissociation equilibrium constant ($K_d$)

$$y = S_f + \frac{(S_b - S_f) \cdot [(x \cdot L \cdot K_d) - ((x \cdot L \cdot K_d)^2 - \sqrt{4 \cdot x \cdot L})]}{2 \cdot L} \tag{2.8}$$

where $y$ = observed fluorescence anisotropy, $x$ = concentration of ligand (M), $K_d$ = dissociation equilibrium constant (M), $S_b$ = maximum observed anisotropy, $S_f$ = observed anisotropy of DNA alone, and $L$ = total concentration of fluorescently-labelled DNA (M).

## 2.12 Crystallisation screening of Cdc45–Treslin complex

The complexes of Cdc45–Treslin$_{626–813}$ or Cdc45$_{\Delta 154–164}$–Treslin$_{626–813}$ at 200 $\mu$M (final concentration) were prepared by reconstituting it over a Superdex75 16/600 size-exclusion column (GE Healthcare) or simply by mixing. 40 $\mu$L of the solution, with 1:1 or 1:2 protein-to-buffer ratio, was used to set up to each commercial screening plate. The screening plates used were JCSG+, Morpheus, PACT, Proplex, and Wizard I & II (Molecular Dimensions). For each condition, duplicates were prepared and each of the plate was incubated either at 19, 11, or 4 °C. After inspecting each of the plate for more than eight weeks, no crystal hit was observed.

No crystal was observed when repeating the experiments using Cdc45–Treslin$_{626–1053}$ or Cdc45$_{\Delta\ 154–164}$–Treslin$_{626–1053}$ complexes, with or without S-CDK phosphorylation treatment.

## 2.13 Sample preparation and data acquisition for cryo-EM

Details for each data acquisition process can be found in the appendix C. Each of the prepared grid was screened in a Talos Arctica electron microscope (FEI ThermoFisher) operated at 200 kV, to judge the particle distribution and ice thickness, before performing image acquisition.

### 2.13.1 Cdc45–Treslin$_{SHD}$ complex

Purified Cdc45 at 110 $\mu$M was incubated with purified Treslin$_{SHD}$ at 550 $\mu$M Treslin$_{SHD}$ for 30 min at 4 °C. The sample was then applied to a Superdex75 10/300 size-exclusion chromatography column (GE Healthcare) equilibrated in *EM Buffer-A* (20 mM HEPES-

NaOH pH 8.0, 100 mM NaCl). The peak fraction — corresponding to the reconstituted complex at 4.9 $\mu$M — was used for cryo-grid preparation. In the case of crosslinked complex, the purified sample was treated with 0.01% (v/v) glutaraldehyde and incubated for 5 min on ice before grid vitrification.

A 3 $\mu$L sample was applied to a R0.6/1.0 UltrAuFoil 300-mesh holey grid (Quantifoil). Prior to sample application, the grid was glow discharged at 25 mA for 1 min each side (PELCO easiGlow). After 2.5 s blotting time at 0 blot force, the grid was vitrified into liquid ethane by using a Vitrobot Mark IV (FEI ThermoFisher) operated at 4 °C and 95% humidity.

The data for each sample were recorded on a Titan Krios electron microscope (FEI ThermoFisher) operated at 300 kV equipped with Volta Phase Plate (VPP) and K2 Summit direct electron detector in counting mode (Gatan Inc.) at the Department of Biochemistry, University of Cambridge. The movies were recorded with a calibrated pixel size of 0.66 Å at a nominal magnification of 215,000 ×. The recordings were divided into 40 frames and automatically collected by EPU software (ThermoFisher) at a dose rate of 1.29 $e^-$ per $Å^2$ per frame. 1,521 movies were collected for non-crosslinked sample, whereas 1,592 for crosslinked sample.

### 2.13.2 Cdc45–Treslin$_{\text{SHD-ext}}$ complex

Purified Treslin$_{\text{SHD-ext}}$ at 35 $\mu$M was phosphorylated by adding 250 nM S-CDK and 1 mM ATP in *EM Buffer-B* (20 mM Tris-HCl pH 7.5, 150 mM K-glutamate, 5 mM $MgCl_2$). The reaction was incubated for 1 h at 25 °C, before adding purified Cdc45 at 70 $\mu$M. A 60 $\mu$L of the reaction sample was subsequently applied to a Superose6 Increase 3.2/300 size-exclusion chromatography column (GE Healthcare) equilibrated in *EM Buffer-B*. The peak fraction at 5 $\mu$M was used for cryo-grid preparation.

A 3 $\mu$L sample was applied to a R0.6/1.0 UltrAuFoil 300-mesh holey grid (Quantifoil). Before applying the sample, the grid was glow discharged at 25 mA for 1 min each side (PELCO easiGlow). After 2.5 s blotting time at -5 blot force, the grid was vitrified into liquid ethane by using a Vitrobot Mark IV (FEI ThermoFisher) operated at 4 °C and 100% humidity.

The data were acquired on a Titan Krios electron microscope operated at 300 kV equipped with Volta Phase Plate and K2 Summit electron detector in 'counting' mode at the eBIC, Diamond Light Source, Oxfordshire. The movies were recorded with a calibrated pixel size of 0.85 Å at magnification of 165,000 ×. The recordings were divided into 35 frames and automatically collected by EPU software at dose rate of 2.32 $e^-$ per $Å^2$ per frame. 1,170 movies were acquired at the end of the session.

### 2.13.3 Apo-Mcm2–7

Purified Mcm2–7 (section 2.3.1) was concentrated to 3–5 mg/mL using Amicon Ultra 0.5 mL centrifugal filter 10 K MWCO (Merck Millipore). A 60 $\mu$L of the concentrated sample was applied to a Superose6 Increase 3.2/300 size-exclusion column (GE Healthcare) equilibrated in *EM Buffer-C* (20 mM Tris-HCl pH 7.5, 50 mM K-glutamate, 5 mM $MgCl_2$). The peak fraction at 0.1 mg/mL was used for further grid preparation.

A 3 $\mu$L sample was applied to a R1.2/1.3 Quantifoil 300-mesh holey copper grid with a thin layer of carbon film (Quantifoil). Before sample application, the grid was glow discharged at 30 mA for 1 min (PELCO easiGlow). After 2.5 s blotting time at -10 blot force, the grid was vitrified into liquid ethane by using a Vitrobot Mark IV (FEI ThermoFisher) operated at 4 °C and 100% humidity.

The data were collected on a Titan Krios electron microscope operated at 300 kV equipped with K2 Summit direct electron detector in 'counting' mode at the eBIC, Diamond Light Source, Oxfordshire. The movies were acquired with a calibrated pixel size of 1.047 Å at nominal magnification 73,000 ×. The recordings were fractionated into 80 frames and automatically collected by EPU software at dose rate 0.66 $e^-$ per $Å^2$ per frame. A total of 3,350 movies were collected.

### 2.13.4 Mcm2–7 + ATP

The grid preparation step was as described for the apo Mcm2–7 sample (section 2.13.3), except that the eluted Mcm2–7 sample after size-exclusion chromatography step was incubated with 2 mM ATP (NEB) for 1 h at 4 °C before applying to the grids.

The data collection was done using a Titan Krios electron microscope operated at 300 kV and equipped with Falcon 3 electron detector in linear mode (FEI ThermoFisher) at the Department of Biochemistry, University of Cambridge, UK. The movies were acquired with a calibrated pixel size of 1.048 Å at nominal magnification 73,000 ×. The recordings were divided into 40 frames and automatically collected by EPU software at dose rate 0.66 $e^-$ per $Å^2$ per frame. A total of 1,350 movies were collected.

### 2.13.5 $\text{Treslin}_{\text{SHD-ext}}$–Mcm2–7 complex

The eluted sample of $\text{Treslin}_{\text{SHD-ext}}$–Mcm2–7 complex originating from the StrepTactin pulldown (section 2.5.5) was isolated and spun at 14,000 *g* for 5 min at 4 °C.

A 3 $\mu$L sample was applied to a R0.6/1.0 UltrAuFoil 300-mesh holey grid (Quantifoil). Prior to sample application, the grid was glow discharged at 25 mA for 1 min each side

(PELCO easiGlow). After 2.5 s blotting time at -5 blot force, the grid was rapidly frozen into liquid ethane by using a Vitrobot Mark IV (FEI ThermoFisher) operated at 4 °C and 95% humidity.

The data were acquired on a Titan Krios electron microscope operated at 300 kV and equipped with K2 Summit direct electron detector in 'counting' mode at the eBIC, Diamond Light Source, Oxfordshire. The data were acquired with a calibrated pixel size of 1.048 Å at nominal magnification of 73,000 ×. The movie recordings were fractionated into 40 frames and collected by EPU software at dose rate 0.88 $e^-$ per $Å^2$ per frame. 3,171 movies were obtained.

## 2.14 Cryo-EM data processing

All the data processing steps were done using RELION software package (Scheres, 2016) installed in the University of Cambridge's High Performance Computing (HPC) facility. All other programs mentioned below were built-in as a package inside RELION software. Different versions of RELION used during the data processing are specified within each subsection. More details can be found on appendix C.

All frames in the recorded movies were corrected for the electron beam-induced motion by using MotionCor 2.1.1, in $5 \times 5$ patch alignment (Zheng et al., 2017). The contrast transfer function (CTF) for each non-dose-weighted micrograph was estimated by using Gctf 1.0.6 (Zhang, 2016). The micrographs with bad ice rings (due to ice contamination) or poor CTF estimation were discarded. When Volta Phase Plate (VPP) was used during image acquisition, the resulting phase shifts were estimated in RELION 3.0 from 0° to 180° with increment every 10°.

### 2.14.1 Cdc45–Treslin$_{SHD}$ complex

#### Non-crosslinked sample

1,326 particles were manually picked, extracted using a box size of 200 pixels, and subjected to a reference-free 2D classification. Five decent class averages were used as template for automatic picking of 393,780 particles. These particles were extracted with a box size of 200 pixels (rescaled by a factor of 4). The particles were subjected to three consecutive rounds of reference-free 2D classification, limiting resolution E-step to 12 Å. The decent looking class averages with total of 78,788 particles were retained.

The data were also re-processed using cryoSPARC 2.8.0 software package (Punjani et al., 2017) installed in a local workstation of the laboratory. 393,210 particles from 1,521

motion-corrected micrographs were automatically picked. These particles were extracted using a box size of 200 pixels and subjected to a reference-free 2D classification. 71,293 particles were retained for ab-initio model generation (asking for 3 classes). A satisfactory ab-initio model was used as a reference for both heterogenous and homogenous refinements.

#### Crosslinked sample

1,445 particles of manually picked particles were extracted with box size of 200 pixels and subjected to reference-free 2D classification. Four decent class averages were used for automatic picking. The resulting 398,556 particles were extracted by using a box size of 200 pixels (rescaled by a factor of 4). The particles were subjected to three consecutive rounds of reference-free 2D classifications, limiting resolution E-step to 12 Å. The well-aligned averages were retained, corresponding to a total of 25,556 particles.

The data were also re-processed using cryoSPARC 2.8.0 software package (Punjani et al., 2017) installed in a local workstation of the laboratory. 339,783 particles from 1,592 motion-corrected micrographs were automatically picked. These particles were extracted using a box size of 200 pixels and subjected to a reference-free 2D classification. 89,134 particles corresponding to 'big' particles (i.e. Cdc45–Treslin$_{SHD}$) were subjected to another round of 2D classification, retaining 50,314 particles. These particles were used for ab-initio model generation (asking for 3 classes). A satisfactory ab-initio model was used as a reference for both heterogenous and homogenous refinement.

### 2.14.2 Cdc45–Treslin$_{SHD\text{-}ext}$ complex

2,450 particles were manually picked, extracted using box size of 180 pixels, and subjected to a reference-free 2D classification. Six satisfactory class averages were used as template for automatic picking of 132,610 particles. These particles were extracted with a box size of 180 pixels (rescaled by a factor of 2). The particles were subjected to two consecutive rounds of reference-free 2D classifications. Particles with badly aligned averages were discarded, retaining 24,827 particles.

To generate an initial 3D model, RELION 3.0's implementation of stochastic gradient descent (SGD) was used. The resulting ab-initio model was low-pass filtered to 50 Å before being used as a template for the 3D classification. Two different classes were produced, but only one class (class 2) with 54.4% of the total particles were selected for further 3D auto-refinement.

### 2.14.3 Apo-Mcm2–7

6,000 particles were manually picked, extracted with box size of 244 pixels, and subjected to reference-free 2D classification. Seven decent class averages were used as template for automatic picking of 644,719 particles. These particles were extracted with a box size of 244 pixels (rescaled by a factor of 2). The extracted particles with rlnParticleSelectZScore more than 0.45 were discarded, while the remaining 552,795 particles were kept and subjected to five rounds of reference-free 2D classifications. Particles with badly-aligned averages were discarded to retain 65,655 particles.

To generate an initial model, RELION 2.1 's implementation of stochastic gradient descent (SGD) was used. A randomly selected subset of 6,000 particles (randomly selected 200 particles $\times$ 30 chosen class averages from the final 2D classifications). The result from the third iteration was low-pass filtered and used as reference for a subsequent 3D classification.

Two different classes were produced, but class 2 with 49.4% of the total particles were selected for further 3D auto-refinement because of the resulting 3D reconstructed map was sharper and not anisotropic. The final random half sets of the dataset converged at a resolution of 6.8 Å as determined by gold standard Fourier shell correlation.

The auto-refined class average was sharpened and its B-factor was estimated by using a solvent mask at binarisation threshold of 0.025 (as inspected from Chimera), 3 pixel edge extension, and raised-cosine soft edge at a width of 3 pixels.

### 2.14.4 Mcm2–7 + ATP

1,006 particles were manually picked, extracted with box size of 244 pixels, and subjected into a reference-free 2D classification. Six decent-looking class averages were used as template for automatic picking of 88,371 particles. These particles were extracted with binning factor of 2 and those with rlnParticleSelectZScore more than 0.45 were discarded. 82,287 particles were kept and subjected to a reference-free 2D classification. Particles with unsatisfactory averages were removed, while keeping the remaining 25,300.

RELION 2.1's method of stochastic gradient descent (SGD) was used to obtain an initial model based on a randomly selected subset of 6,000 particles. The result from a third iteration was used as reference for the subsequent 3D classification after low-pass filtering.

Two different classes were produced, but only class 1 with 49.6% of the total particles were selected for further 3D auto-refinement. The final random half sets of the dataset converged at a resolution of 10.1 Å as determined by gold standard Fourier shell correlation.

The auto-refined class average was sharpened and its B-factor was estimated by using a solvent mask at binarisation threshold of 0.0305 (as inspected from Chimera), 3 pixel edge extension, and raised-cosine soft edge at a width of 5 pixels.

### 2.14.5 $\text{Treslin}_{\text{SHD-ext}}$–Mcm2–7 complex

1,189 particles were manually picked, extracted with box size 280 pixels, and subjected to a reference-free 2D classification. Seven satisfactory class averages were selected and used as templates for automatic picking in all the micrographs. 533,326 resulting particles were automatically picked and extracted with pixel size of 280 (rescaled with a factor of 4). The extracted particles were sorted using their rlnParticleSelectZScore and those with value less than 0.5 were discarded, keeping only 490,303 particles. The remaining particles were subjected to two consecutive reference-free 2D classifications. 118,706 particles were then saved from satisfactory well-aligned class averages were retained.

RELION 3.0's method of stochastic gradient descent (SGD) was utilised to obtain an initial model based on the selected particles. This was low-pass filtered to 50 Å and used as a reference model for the subsequent 3D classifications in RELION 3.0, asking for 3 classes. One class, class1, was reconstructed from 37.3% of the total particles (44,300 particles). The corresponding particles were re-extracted using box size of 360 pixel (without re-scaling) and subjected to 3D auto-refinement. The final random half sets of the dataset converged to a resolution of 8.2 Å as determined by gold standard Fourier shell correlation in the RELION 3.0. Unfortunately, there was a strong evidence of preferred orientation (anisotropic view) in the 3D reconstructed map.

The auto-refined class average was sharpened and its B-factor was estimated by using a solvent mask at binarisation threshold of 0.0075 (inspected in Chimera), 5 pixel edge extension, and raised-cosine soft edge at a width of 6 pixels.

## 2.15 SDS-PAGE

A 10 $\mu$L of each sample for SDS-PAGE analysis (Sodium Dodecyl Sulphate - Polyacrylamide Gel Electrophoresis) was prepared by adding 1.5 $\mu$L of 10$\times$ NuPAGE reducing agent (Invitrogen) and 3.5 $\mu$L of 4$\times$ NuPAGE LDS sample buffer (Invitrogen). The sample was heated at 95 °C for 1 min before loading on the SDS-PAGE gel. SeeBlue or SeeBlue Plus 2 protein markers (Invitrogen) were also loaded. The gel was run using 1$\times$ NuPAGE MOPS-SDS buffer (Invitrogen) at 200 V for 40 min (Mini Gel Tank Invitrogen and BioRad PowerPac Basic). The gel was microwaved in distilled water for 30 s, stained for 2 $\times$ 5 min

with Coomassie Blue Stain (Invitrogen), and destained by washing it repeatedly with distilled water. In case of silver-staining gel, the Pierce$^{TM}$ silver stain kit (ThermoFisher) was used and the protocol was followed as recommended by the manufacturer.

The 12% Bis-Tris SDS-PAGE gel was set in 1 mm Bolt Cassettes (Invitrogen) by pouring separating gel mixture consisting of 12 % acrylamide (37.5:1; 30% acrylamide), 0.36 M Bis-Tris pH 6.8, 0.0011% (v/v) of tetramethylethylenediamine (TEMED) and 0.035% (v/v) ammonium persulfate (APS). After the separating gel had polymerised, the stacking gel mixture consisting of 4% acrylamide (40% acrylamide), 0.36 M Bis-Tris pH 6.8, 0.0011% (v/v) of TEMED and 0.035% (v/v) APS was poured and left to set. The gel was kept wet at 4 °C.

In case of a gradient gel analysis, a 4–12% pre-cast Novex SDS-PAGE gel (Invitrogen) was used.

## 2.16 Western blot

Protein samples were separated by SDS-PAGE (refer to section 2.15). The resulting gel was blotted onto a Novex 0.45 $\mu$m nitrocellulose membrane (Invitrogen) in 1× NuPAGE transfer buffer (Invitrogen) + 10% (v/v) methanol, by using a Novex transfer cassette (Invitrogen) at 15 V for 2 h.

### Protein detection

After transfer, the membrane was blocked by incubating it with 5 mL of 5% (w/v) milk dissolved in 1× PBST buffer (PBS + 0.2% Tween 20) at roller mixer for 10 min at 4 °C. The primary antibody was added and the membrane was further incubated on a roller mixer for 2 h at room temperature or overnight at 4 °C.

The membrane was washed thrice with 20 mL of 1× PBST buffer for 5 min each time at room temperature to remove the excess unbound antibody. The membrane was blocked again, as described previously, before the addition of secondary antibody. The membrane was incubated at roller mixer for 1 h at 4 °C.

The membrane was washed as described previously. The bands on the membrane were detected using Amersham ECL Western Blot Detection Reagent kit (GE Healthcare).

The list of the antibodies used can be found in the following table.

| Antibody | Working dilution | Manufacturer (catalogue number) |
|---|---|---|
| mouse monoclonal IgG anti-Cdc45 | 1:5000 | SantaCruz (sc-55569) |
| mouse monoclonal IgG anti-Mcm7 | 1:5000 | SantaCruz (sc-9966) |
| mouse monoclonal IgG anti-polyHis tag | 1:5000 | Sigma Aldrich (H1029) |
| rabbit monoclonal IgG anti-TwinStrep tag | 1:1000 | RayBiotech (168-10750) |
| rabbit monoclonal IgG anti Mcm2-pSer40/41 | 1:5000 | (Montagnoli et al., 2006) |
| anti-mouse IgG-HRP conjugate* | 1:5000 | Promega (W4021) |
| anti-rabbit IgG-HRP conjugate* | 1:5000 | Promega (W4011) |

* = secondary antibody for HRP detection.

### Biotinylation detection

After transfer, the membrane was blocked by incubating it with 5 mL of 5% (w/v) milk dissolved in 1× PBST buffer (PBS + 0.2% Tween 20) on a roller mixer for 10 min at 4 °C. Streptavidin conjugated to alkaline phosphatase (Roche) in 1:5000 working dilution was added to the solution and left on a roller mixer for 3 h at 4 °C.

The membrane was washed 3 times with 20 mL of 1× PBST buffer at room temperature. The biotinylated bands on the membrane were detected by using a tablet of BCIP/NBT (5-bromo-4-chloro-3-indolyl phosphate / nitro blue tetrazolium) (Sigma) dissolved in 10 mL of deionised water.

# Chapter 3

# Cdc45–Treslin interaction

## 3.1 Rationale

Cdc45 (Cell Division Cycle 45 protein) is an integral component of eukaryotic replisome that is required throughout the DNA replication process. Specifically, it is a constituent of the CMG holo-helicase complex that unwinds the parental dsDNA to provide ssDNA templates for the replicative DNA polymerases.

Recruitment of Cdc45 to the DNA replication origin occurs during the G1 phase of cell cycle and is one of the limiting steps of DNA replication initiation (Kamimura et al., 2001; Tanaka et al., 2011a). In *S. cerevisiae*, this process is dependent on phosphorylation of MCM by DDK and the action of Sld3 protein (Deegan et al., 2016; Heller et al., 2011; Yabuuchi et al., 2006). The role of Sld3 is proposed to be a loading factor of Cdc45 to the DNA replication origin-bound MCM, where Sld3's interaction with Cdc45 is mediated by a conserved SHD domain (Itou et al., 2014; Sanchez-Pulido et al., 2010). In higher eukaryotes, Treslin was identified as the functional orthologue of Sld3. However, its molecular details — especially how Treslin binds to Cdc45 — are largely unclear (Kumagai et al., 2010, 2011; Sanchez-Pulido et al., 2010; Sansam et al., 2010).

In this chapter, the Cdc45–Treslin interaction was characterised using purified recombinant human proteins. Previous works in the Pellegrini Lab produced high-resolution structures of Cdc45 and Treslin's SHD domain ($\text{Treslin}_{\text{SHD}}$) individually through X-ray crystallography (Simon et al., 2016). However, the structure of Cdc45–Treslin has yet to be solved.

Cdc45 and $\text{Treslin}_{\text{SHD}}$ were purified and used for pulldown and semi-analytical size-exclusion chromatography experiments. The oligomerisation state of Cdc45–$\text{Treslin}_{\text{SHD}}$ complex was determined by a size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). The binding affinity and the kinetics of the Cdc45–$\text{Treslin}_{\text{SHD}}$

interaction were also measured by Surface Plasmon Resonance (SPR) and Bio-Layer Interferometry (BLI).

This chapter also discusses the biochemical and biophysical characterisations of another Treslin domain called 'extension', hereby referred to as Treslin$_{ext}$. This domain, which is located immediately in the C-terminus of the SHD, is strongly conserved across the vertebrates. Treslin$_{ext}$ contains many conserved positively-charged amino acid residues and potential S-phase Cyclin-Dependent Kinase (S-CDK) phosphorylation sites (threonine/serine-proline or T/S-P motifs) (chapter 1, figure 1.11). Treslin$_{ext}$ interaction with Cdc45 and the possible regulatory role by S-CDK are also discussed.

Lastly, this chapter also describes the latest progress in structural analysis of Cdc45–Treslin complex through single-particle cryo-electron microscopy (cryo-EM) analysis. Low-resolution 3D reconstructed maps of Cdc45–Treslin — Cdc45–Treslin$_{SHD}$ and Cdc45–Treslin$_{SHD\text{-}ext}$ — are presented.

The findings presented in this chapter contribute to a more detailed understanding of Cdc45–Treslin complex's role in the initiation of human DNA replication.

## 3.2 Results and Discussions

### 3.2.1 Treslin$_{SHD}$ interacts with Cdc45

In *S. cerevisiae*, Cdc45 associates with Sld3 prior to its recruitment to the DNA replication origin (Kamimura et al., 2001). This Cdc45-binding function of Sld3 was demonstrated to be conserved in higher eukaryotes, based on studies in *X. laevis* (Kumagai et al., 2011). Genetic screening and bioinformatic analysis subsequently revealed that Sld3 and Treslin share a conserved domain called Sld3/Treslin Homology Domain (chapter 1, figure 1.11) (Kumagai et al., 2011; Sanchez-Pulido et al., 2010). Studies using purified budding yeast proteins subsequently showed that the SHD is required for Cdc45-binding activity, although this has not been verified using purified human proteins (Itou et al., 2014). Thus, Cdc45-binding experiments for human Treslin$_{SHD}$ were performed.

Two constructs representing Treslin$_{SHD}$ were initially cloned. One comprised Treslin residues 626–813 and the other is 503–934, spanning a longer SHD. Both were cloned in the pGAT2 vector to enable the addition of His-GST tags to the N-terminus and overexpression in *E. coli*. After size-exclusion chromatography step, there was less degradation products for Treslin residues 626–813 compared to 503–934, as judged by Coomasie-stained SDS-PAGE. Treslin residues 626–813 was also more readily expressed. Taken together, this indicated that Treslin residues 626–813 behaved more stably (figure 3.1a and b). Both

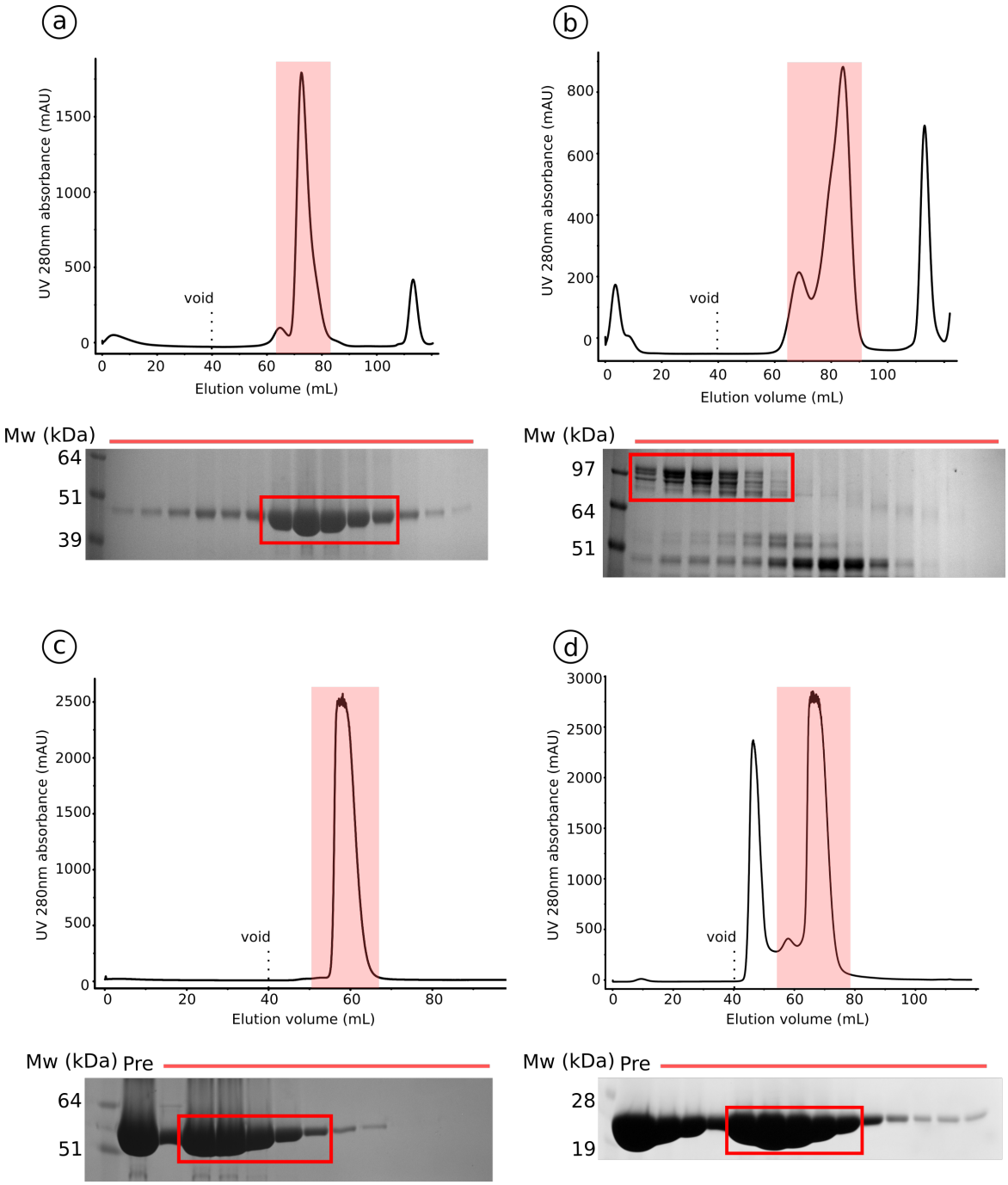



**Fig. 3.1 Treslin$_{SHD}$ and Cdc45 purification.** Coomassie-stained SDS-PAGE showing the fractions from the Superdex75 16/600 size-exclusion chromatography step for purification of **(a)** His-GST tagged Treslin$_{626–813}$, **(b)** His-GST tagged Treslin$_{503–934}$, **(c)** untagged Cdc45, and **(d)** untagged Treslin$_{626–813}$. Red boxes indicate fractions that were pooled for subsequent experiments. Pre = sample before size-exclusion chromatography run.

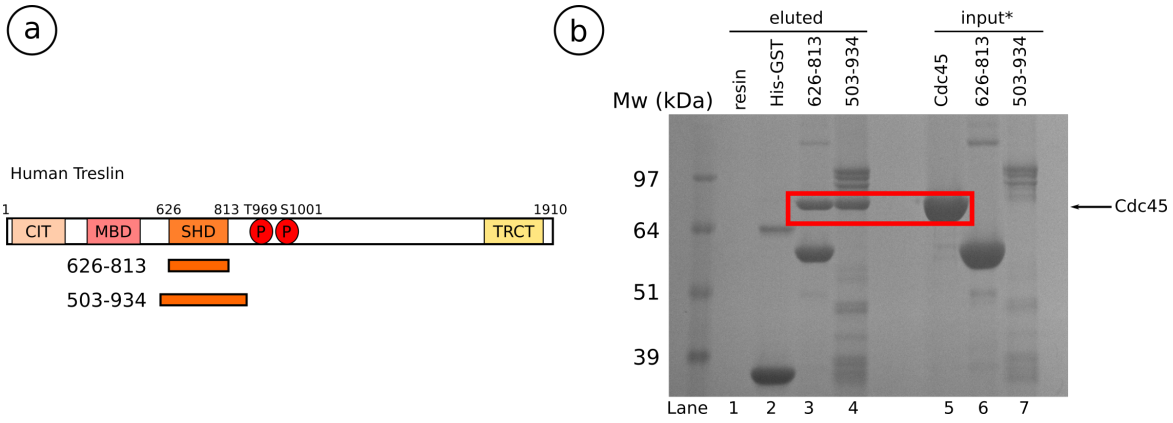



**Fig. 3.2 GST pulldown experiment of His-GST tagged Treslin$_{SHD}$ and Cdc45.** **(a)** His-GST tagged constructs used for this experiment. CIT = Conserved In Treslin, MBD = MTBP-Binding Domain, SHD = Sld3/Treslin Homology Domain, P = conserved S-CDK phosphorylation site, TRCT = Treslin C-Terminal domain. **(b)** Coomassie-stained SDS-PAGE showing the eluted fractions of GST pulldown of various His-GST tagged Treslin constructs with Cdc45. Cdc45 is indicated in red box. The input amount for His-GST-Treslin$_{503–934}$ (lane 7) is 10% of the other proteins, due to limited amount of purified proteins.

His-GST-tagged constructs could be purified over a Superdex200 16/600 size-exclusion chromatography column (GE Healthcare) and subsequently used for pulldown experiments. Meanwhile, human Cdc45 was very readily expressed in *E. coli* and it was possible to purify it to homogeneity after a Superdex75 16/600 size-exclusion chromatography (figure 3.1c).

To verify whether Treslin$_{SHD}$ interacts with Cdc45, a GST pulldown experiment was performed. Purified His-GST tagged Treslin (figure 3.2a) was used as the 'bait' and attached to a glutathione resin. In turn, an excess amount of purified untagged Cdc45 was used as the target protein. The bound proteins were eluted by buffer containing 15 mM reduced glutathione, pH 7.5. As shown in figure 3.2b, Cdc45 was co-eluted from the beads with both constructs of His-GST-Treslin$_{SHD}$. Cdc45 neither co-eluted with free His-GST nor bound non-specifically to the resin. There was no apparent difference in binding ability for both Treslin constructs as judged by Coomassie-stained SDS-PAGE gel (figure 3.2b). This result confirms that Treslin$_{SHD}$ is the conserved region in human Treslin that mediates interaction with Cdc45.

In subsequent experiments, Treslin residues 626–813 rather than Treslin 503–934 was used for two reasons: it bound to Cdc45 as well as the other construct and it expressed very readily. Untagged Treslin$_{SHD}$ (residues 626–813), after removal of the His-tag with TEV protease, could be purified using a Superdex75 16/600 size-exclusion chromatography (figure 3.1d).

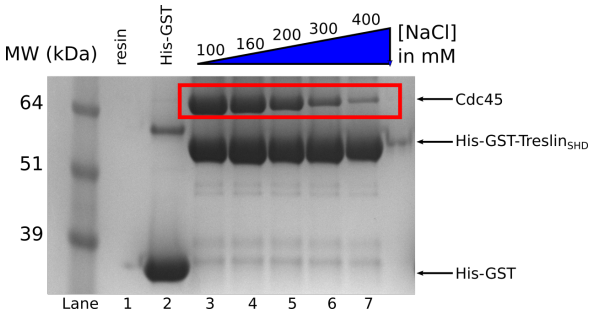



**Fig. 3.3 The effect of salt concentration on Cdc45-Treslin$_{SHD}$ complex formation.** Coomassie-stained SDS-PAGE showing the eluted fractions from GST pulldown of His-GST tagged Treslin$_{SHD}$ (residues 626–813) with Cdc45 in different salt concentrations. Co-eluted Cdc45 is highlighted in red.

### 3.2.2 The Cdc45–Treslin$_{SHD}$ interaction is mediated by electrostatic interactions

A study of the *S. cerevisiae* Cdc45–Sld3 complex suggested that its interaction is driven by electrostatic forces. Here, the acidic surface of Cdc45 was shown to bind to the basic surface of Sld3 (Itou et al., 2014). To test whether this is also true in human system, a GST pulldown experiment of Treslin$_{SHD}$ and Cdc45 was performed in a range of sodium chloride (NaCl) concentration. The bound proteins were eluted from the beads by buffer containing 15 mM reduced glutathione, pH 7.5. The result showed an inverse relationship between increasing NaCl concentration and the amount of Cdc45 co-eluted with His-GST tagged Treslin$_{SHD}$. As the concentration of salt increased, there was less Cdc45–Treslin complex being formed (figure 3.3). This result suggested that Treslin and Cdc45 binding is driven at least partly by electrostatic interactions, similar to the *S. cerevisiae*'s system.

Cdc45–Treslin$_{SHD}$ complex formation was further verified by a semi-analytical gel filtration experiment. Cdc45 was mixed with a 5-fold molar excess of untagged Treslin$_{SHD}$ and applied to a Superdex75 10/300 semi-analytical size-exclusion chromatography column. The complex stayed together over gel filtration as the complex co-migrated and eluted earlier than their individual protein components, indicating formation of a bigger molecular species of Cdc45–Treslin$_{SHD}$ (figure 3.4a and d). When the NaCl concentration in the buffer was increased to 400 mM, the complex appeared to fall apart as demonstrated by less amount of Treslin$_{SHD}$ that co-migrated with Cdc45 (figure 3.4a and e). This result was in agreement with the GST pulldown experiment shown in figure 3.3, where the interaction is partially driven by electrostatic forces and was only stable in low salt conditions.

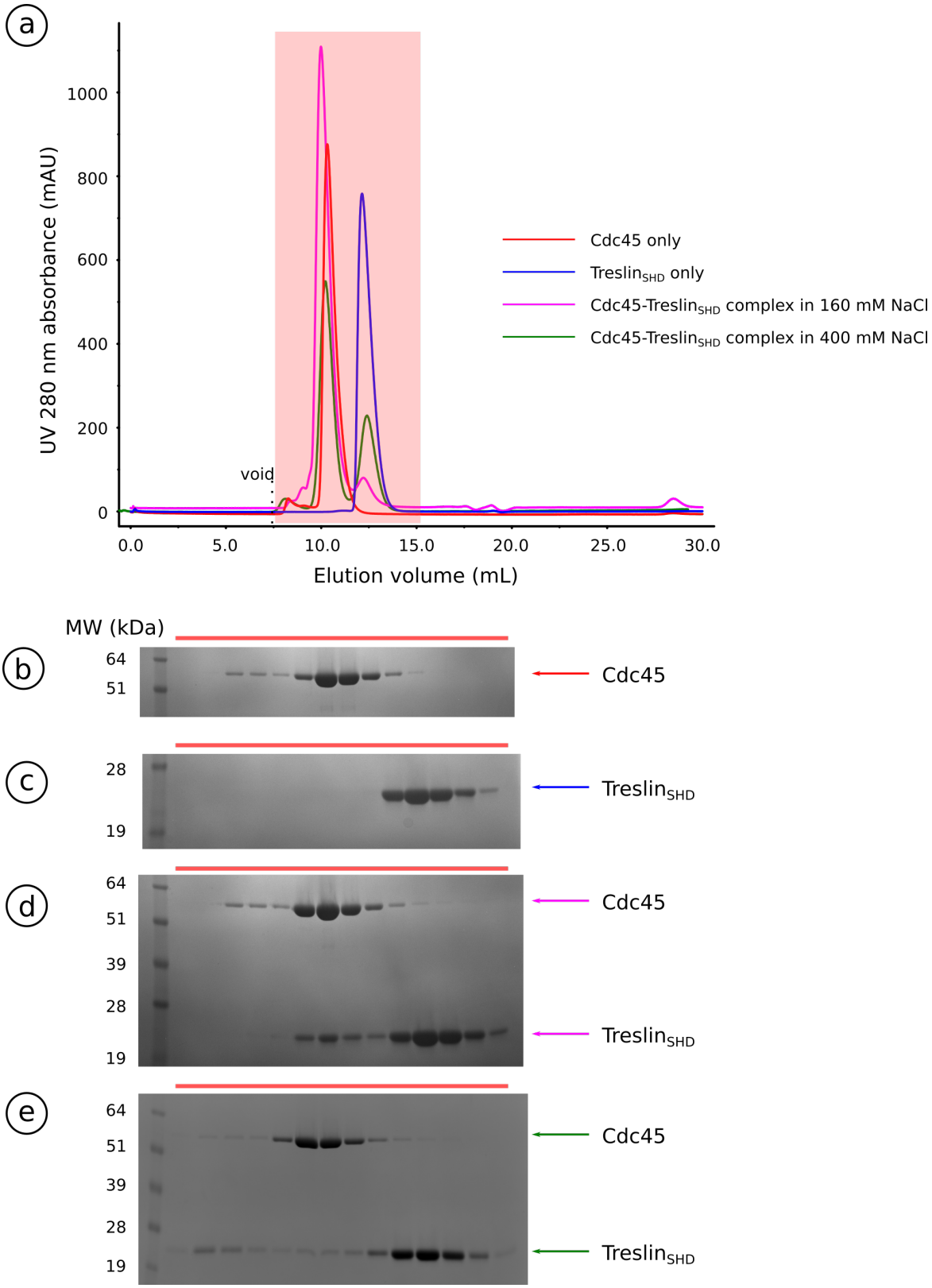



**Fig. 3.4 Semi-analytical gel filtration of Cdc45–Treslin$_{SHD}$ in low or high salt concentration.** **(a)** Chromatogram showing the elution profile in a Superdex75 10/300 size-exclusion chromatography column for Cdc45, Treslin$_{SHD}$, or in complex. Coomassie-stained SDS-PAGE showing the fractions of **(b)** Cdc45 only, **(c)** Treslin$_{SHD}$ only, **(d)** Cdc45–Treslin$_{SHD}$ in 160 mM NaCl, and **(e)** Cdc45–Treslin$_{SHD}$ in 400 mM NaCl.

### 3.2.3 Cdc45–Treslin$_{SHD}$ forms a 1-to-1 complex in solution

The oligomeric state of Cdc45–Treslin$_{SHD}$ was determined by Size-Exclusion Chromatography coupled to Multi-Angle Light Scattering (SEC-MALS). It is an excellent tool to determine the oligomeric state of biomolecules complex in solution. SEC separates the proteins based on their hydrodynamic volume, whereas MALS determines the molecular weight based on the light scattering detected at several angles and protein concentration information obtained from the in-line differential refractometer (chapter 2, section 2.6).

As controls, each of the individual proteins were applied to a Superdex200 Increase 10/300 gel filtration column (GE Healthcare). A 130 $\mu$M Cdc45 sample was applied to the column and eluted in a single peak of 59.8 kDa ($\pm$0.05 kDa), close to its predicted monomeric molecular weight of 65.6 kDa (figure 3.5a). This result showed that Cdc45 exists as a monomer in solution. Next, a 650 $\mu$M Treslin$_{SHD}$ sample — a 5-fold molar excess compared to Cdc45 — was applied to the column. It eluted as a single peak of 20.6 kDa ($\pm$0.04 kDa), close to its expected monomeric molecular weight of 21.5 kDa (figure 3.5b). This observation indicated that Treslin$_{SHD}$ is monomeric in solution.

The two proteins were mixed in 1:5 molar ratio of Cdc45 to Treslin$_{SHD}$ and analysed by SEC-MALS. Two peaks were observed, in which the peak eluting later corresponded to the excess Treslin$_{SHD}$ that did not participate in the complex formation (it was eluted as a monomer of 21 kDa). The peak eluting earlier, displayed a molecular weight of 78 kDa ($\pm$0.1 kDa), 9.1 kDa less than the predicted 1-to-1 complex molecular weight value of 87.1 kDa (addition of 21.5 kDa from monomeric Treslin$_{SHD}$ + 65.6 kDa from monomeric Cdc45, figure 3.5c). The approximately 10 kDa difference between the observed and expected value could be caused by the inability of the Superdex200 Increase 10/300 column to fully resolve the complex and individual Cdc45 completely or by non-specific interaction of the proteins to the column resin. Nevertheless, there was a clear shift in elution volume by 0.5 mL between Cdc45 by itself and the '78-kDa complex' (figure 3.5a vs c). Therefore, this observation indicates that Cdc45–Treslin$_{SHD}$ forming a 1-to-1 complex in solution.

### 3.2.4 Kinetics and binding affinity of the Cdc45–Treslin$_{SHD}$ complex

#### Bio-Layer Interferometry (BLI) measurement

There are currently no kinetic or binding affinity measurement data for Cdc45–Treslin complex — or *S. cerevisiae* Cdc45–Sld3 complex — available in the literature. This information may provide important insights on Cdc45–Treslin complex biochemical behaviour. First, Bio-Layer Interferometry (BLI) measurements were performed using the Octet RED96 system (Pall ForteBio). BLI requires immobilisation of a tagged or labelled protein ligand to

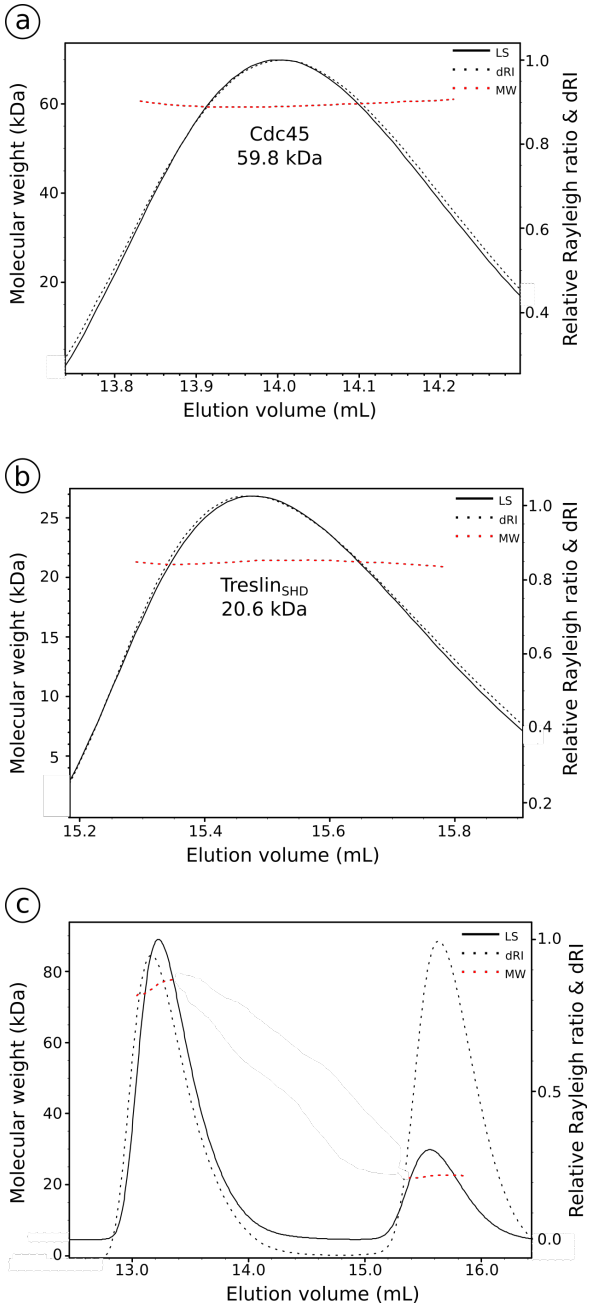



**Fig. 3.5 SEC-MALS of Cdc45–Treslin$_{SHD}$ complex.** **(a)** SEC-MALS result for Cdc45. Polydispersity = 1.001 (±0.089 %). Predicted Mw = 64.1 kDa. **(b)** SEC-MALS result for Treslin$_{SHD}$ (residues 626–813). Polydispersity = 1.000 (±0.277 %). Predicted Mw = 21.5 kDa. **(c)** SEC-MALS result for Cdc45–Treslin$_{SHD}$ complex in 1:5 molar ratio. The complex was eluted in an earlier peak with Mw = 78 kDa. Polydispersity = 1.000 (±0.177 %). Predicted Mw of 1-to-1 complex = 87.1 kDa. The later peak contained the excess Treslin$_{SHD}$. LS = Light scattering, Mw = Molecular weight, dRI = differential Refractometer index.

the tips of the biosensors. The biosensors were used to perform 'dip-and-read' experiments in a 96-well plate containing various dilutions of analytes, which then gave a signal in the change of layer thickness (in nm) at the tip of the biosensor. From this signal read-out, the graph of concentration against the maximum response was plotted to derive the binding affinity ($K_d$, or dissociation constant). Furthermore, the rate of association and dissociation could also be deduced to obtain the kinetics of complex formation from the mean of the dissociation rate obtained (chapter 2, section 2.8).

For immobilisation step, $Treslin_{SHD}$ with an AviTag$^{TM}$ was used. Initially, the construct was co-expressed in *E. coli* with BirA ligase in media containing biotin to promote *in vivo* biotinylation. The biotinylation of $Treslin_{SHD}$ was verified by liquid chromatography mass spectrometry (LC-MS) performed by Dr. Mike Deery at the Cambridge Centre of Proteomics and by western blot using streptavidin-AP conjugated antibodies (Roche). The resulting biotinylated protein was subsequently used for BLI and SPR measurements.

A 10 nM sample of biotinylated Avi-tagged $Treslin_{SHD}$ protein was used to saturate the binding surfaces of the biosensors. The sensors were then immersed in wells containing buffer to remove excess ligand, then into new buffer wells for baseline recording, and finally into the well containing various concentrations of purified untagged Cdc45. For the dissociation step, the sensors were dipped into the same wells as used for the baseline recording. A representative sensorgram is shown in figure 3.6a. Measurements were repeated twice. Based on the results of steady-state analysis, the mean $K_d$ of Cdc45–$Treslin_{SHD}$ was 4.8 $\mu$M (figure 3.6b).

The dissociation rate ($K_{off}$) of Cdc45 and $Treslin_{SHD}$ complex was determined as 0.922 $s^{-1}$, which indicated an extremely fast rate of complex disassembly. The division of $K_d$ with the average $K_{off}$ value gave an association rate ($K_{on}$) of $1.92 \times 10^5$ $M^{-1}$ $s^{-1}$. These association and dissociation rates were surprising, as they appeared to be very fast for a protein complex of Cdc45–$Treslin_{SHD}$ with molecular weight of approximately 80 kDa. These values would be expected for small molecules or peptides interactions. Nevertheless, they were still within the biological limits for protein–protein interaction and solvent diffusion limit (Schlosshauer and Baker, 2004). There was no indication of non-specific interaction of the analyte to the biosensors and the data fitting was satisfactory (figure 3.6a).

### Surface Plasmon Resonance (SPR) measurement

Surface Plasmon Resonance (SPR) is another technique, with a different principle from BLI, which was utilised to measure the binding affinity and kinetics of Cdc45–$Treslin_{SHD}$ complex. This technique relies on an optical phenomenon termed Surface Plasmon Resonance. SPR, like BLI, also requires an immobilisation of the protein ligand. The immobilised ligand is

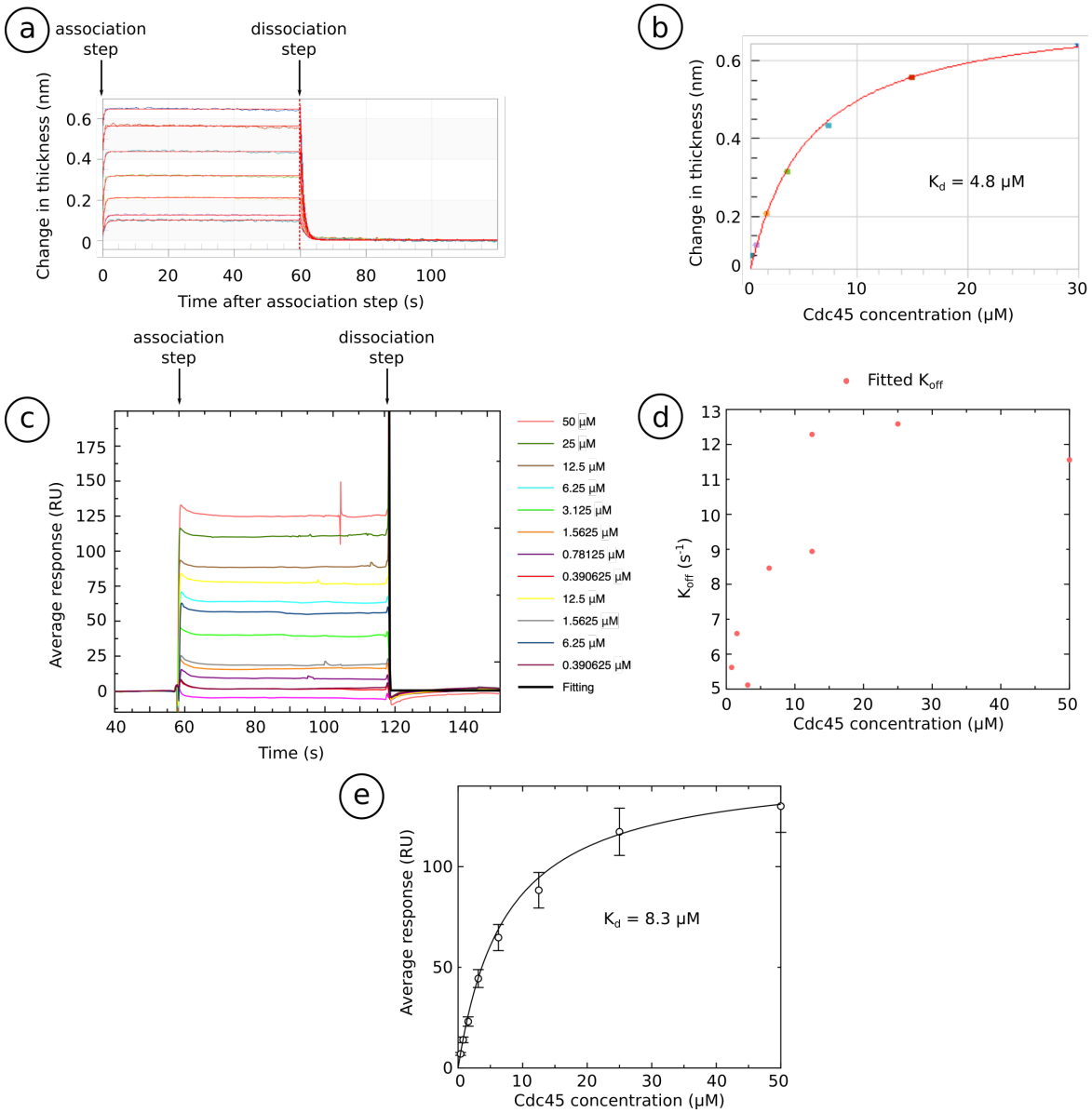



**Fig. 3.6 BLI and SPR measurements of Cdc45–Treslin$_{SHD}$ complex.** **(a)** Sensorgram of Bio-Layer Interferometry (BLI) measurement with 1:1 global fitting model. **(b)** Steady state analysis using the results from the BLI measurements. **(c)** Sensorgram of Surface Plasmon Resonance (SPR) measurement with first order decay fitting model. **(d)** Plot of fitted dissociation rate ($K_{off}$) against Cdc45 concentration. Average $K_{off}$ = 8.89 $s^{-1}$. **(e)** Steady state analysis using the results from the SPR measurements.

then subjected to a continuous flow of buffer which can be set to include protein analytes in various concentrations (chapter 2, section 2.9). Thus, the SPR experiment was performed to verify the results obtained from BLI. A 100 nM sample of biotinylated Avi-tagged $Treslin_{SHD}$ was used for immobilisation on a two-dimensional, gold-layered sensor chip containing streptavidin (XanTec Bioanalytics).

Steady state analysis was obtained by plotting and fitting the analyte concentration against the equilibrium response observed from the sensorgram. The mean $K_d$ from three separate measurements for Cdc45–$Treslin_{SHD}$ based on SPR was 8.3 $\mu$M. This result was similar and within the same order of magnitude as the value observed from BLI. This very fast kinetics of the complex were confirmed by SPR (figure 3.6c). It needs to be noted that there was no clear linear trend in the plot of $K_{off}$ against Cdc45 concentration, therefore the $K_{off}$ values were averaged (mean $K_{off}$ = 8.9 $s^{-1}$) (figure 3.6d). The average $K_{off}$ and the $K_d$ obtained from the steady state analysis was divided to derive the $K_{on}$ = 1.069 $\times$ $10^6$ $M^1$ $s^{-1}$. In conclusion from two different biophysical measurements, binding affinity ($K_d$) of Cdc45–$Treslin_{SHD}$ complex were resulting in similar binding constants with values between 4.8–8.3 $\mu$M.

In addition, the Cdc45–$Treslin_{SHD}$ complex has a very rapid rate of association and dissociation as indicated by their $K_{on}$ and $K_{off}$ values. One speculation is that the fast kinetics is a way to ensure rapid recycling of Treslin for late DNA replication origins. In *S. cerevisiae*, the Sld3 protein levels are low throughout cell cycle and represent a limiting factor for the Cdc45-recruitment step (Mantiero et al., 2011).

### 3.2.5 $Treslin_{ext}$ is intrinsically disordered and a substrate for S-CDK

A sequence homology analysis of the C-terminal region immediately after the $Treslin_{SHD}$ revealed a high degree of conservation in higher eukaryotes (figure 3.7a). This particular domain shall be referred to as 'extension' domain ($Treslin_{ext}$). This region contains a number of highly-conserved positively-charged residues. It also contains at least six potential S-CDK phosphorylation sites, including two putative sites for TopBP1 binding (T969 and S1001) (Boos et al., 2011; Errico et al., 2010; Kumagai et al., 2011). Currently, there is no information whether $Treslin_{ext}$ is capable of binding to Cdc45 and whether S-CDK sites are involved in the regulation.

Freely available secondary structure prediction softwares such as IUPred and ANCHOR are useful to predict whether $Treslin_{ext}$'s sequence is intrinsically disordered or able to bind to other proteins. IUPred's algorithm estimates pairwise energy content that assumed globular proteins consisted of amino acids that may form large number of favourable interactions. Meanwhile, ANCHOR predicts segments in the disordered region that may gain stabilising energy and fold properly upon binding to another globular protein. ANCHOR's result relies

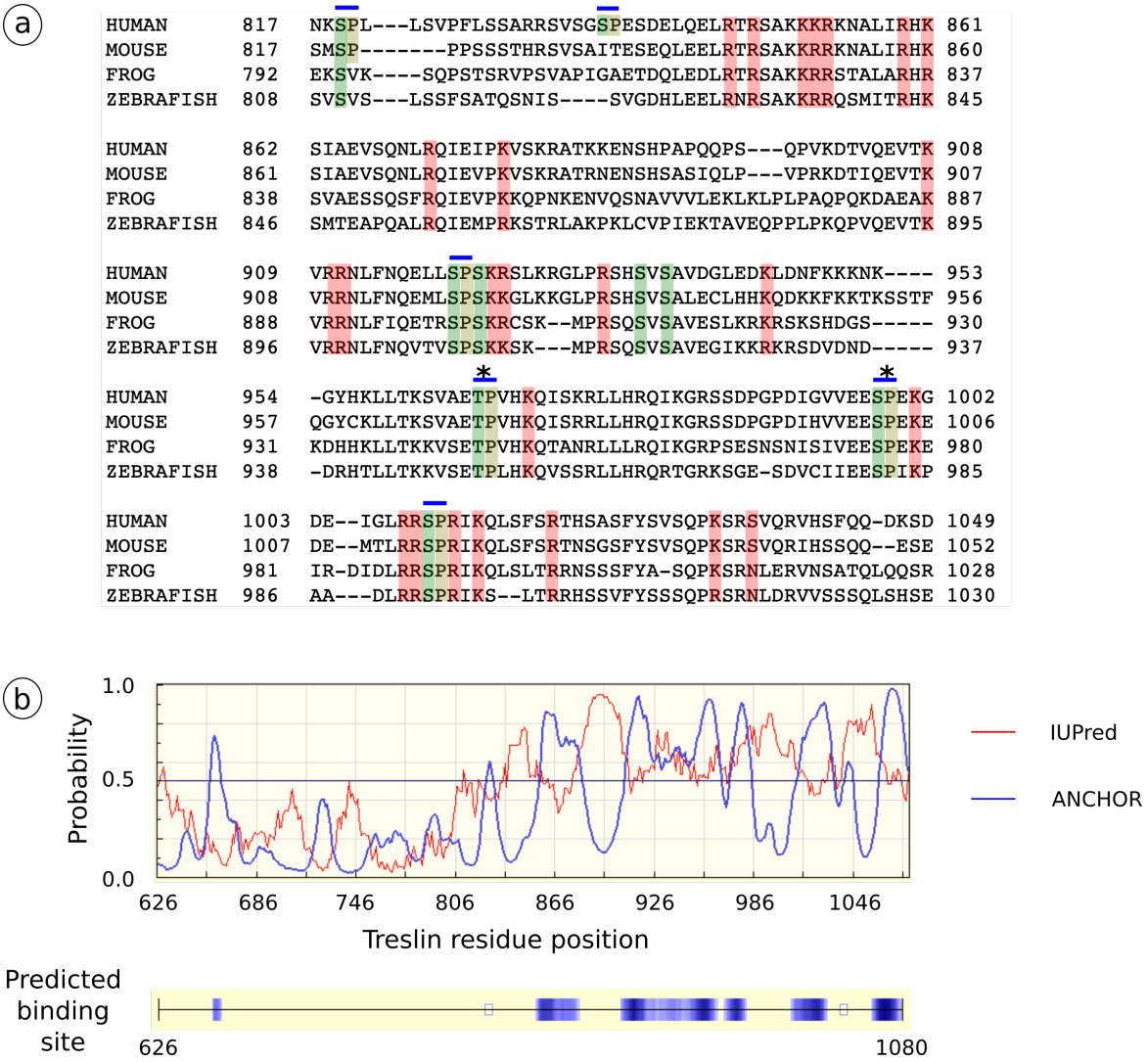



**Fig. 3.7 Sequence conservation of Treslin$_{ext}$ (a)** Sequence conservation of Treslin$_{ext}$ domain (residue 817–1049) across higher eukaryotes. The potential S-CDK sites are marked with blue line. The putative sites for TopBP1 binding are marked with asterisk (Boos et al., 2011; Kumagai et al., 2011). Conserved positively charged residues = red, polar residues = green, and non-polar proline residues = olive. **(b)** Treslin$_{SHD\text{-}ext}$ disordered region prediction from residue 626 to 1080. Red line (IUPred) indicates the probability of that particular region folds into disordered region. Blue line (ANCHOR) shows the probability that this region folds upon binding of interactors (Dosztanyi et al., 2009; Meszaros et al., 2009).

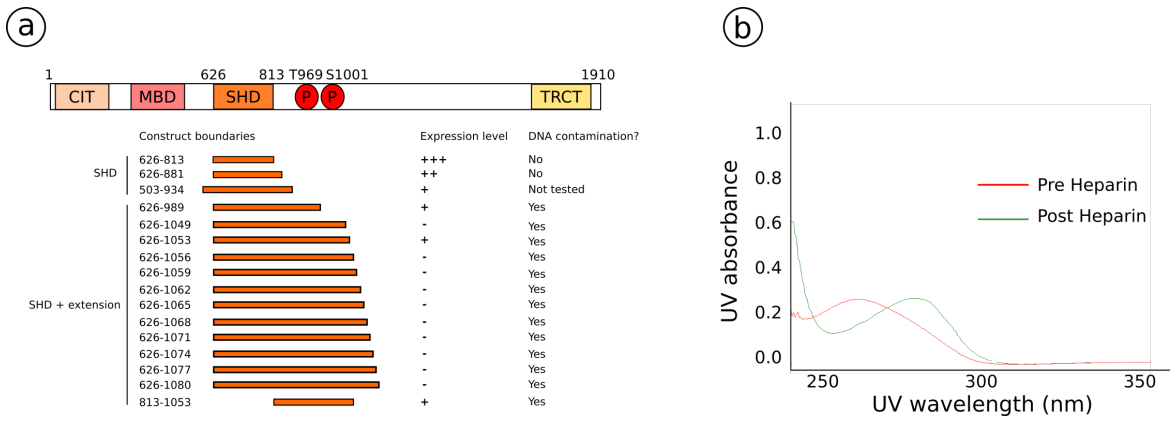



**Fig. 3.8 Construct screening for Treslin$_{\text{SHD-ext}}$.** **(a)** Overview of the Treslin$_{\text{SHD-ext}}$ constructs screening. Residue numbers are indicated on the left. +++ = very good expression level of soluble proteins, ++ = good expression level of soluble proteins, + = low expression level of soluble proteins, and - = very low or no expression level of soluble proteins. CIT = Conserved in Treslin, MBD = MTBP-Binding Domain, SHD = Sld3/Treslin Homology Domain, P = Conserved Phosphorylation Sites, TRCT = Treslin C-Terminal. **(b)** UV absorbance reading for Treslin$_{\text{SHD-ext}}$ before and after Heparin-affinity column chromatography.

on pairwise energy estimation in IUPred (Dosztanyi et al., 2009; Meszaros et al., 2009). Treslin$_{\text{ext}}$ is predicted to be highly disordered as judged by the IUPred score. The ANCHOR prediction also suggested that Treslin$_{\text{ext}}$ may fold upon binding to globular proteins (figure 3.7b, highlighted in blue blocks), suggesting that this region contains binding sites of other protein partners. One candidate interactor is Cdc45, which is discussed in section 3.2.6.

A series of constructs representing Treslin$_{\text{SHD-ext}}$ (Treslin SHD + 'extension' domain) and Treslin$_{\text{ext}}$ were generated and screened by truncating every 3 amino acids from the C-terminus of the extension domain (figure 3.8a). Purification of Treslin$_{\text{SHD-ext}}$ proved to be challenging due to extremely low expression level, C-terminus degradation, poor solubility, and DNA contamination (figure 3.8a). These problems are likely to be caused by the disordered nature of the Treslin$_{\text{ext}}$ as predicted earlier (figure 3.7b). These challenges could partly be alleviated by expressing a construct with an N-terminal MBP tag which slightly improve the solubility and engineering C-terminal octahistidine tag in order to separate it from the C-terminal degradation products, and increasing the *E. coli* culture volume to at least 12 litres. DNA contamination persisted in the sample despite benzonase addition and/or high salt washes, and could only be separated from the protein-of-interest by using Heparin-affinity column chromatography (figure 3.8b).

Treslin constructs corresponding to residues 626–1053 and 813–1053 were selected for further experiments, representing Treslin$_{\text{SHD-ext}}$ and Treslin$_{\text{ext}}$, respectively, because they were expressed better than other constructs. Purification steps of these two constructs

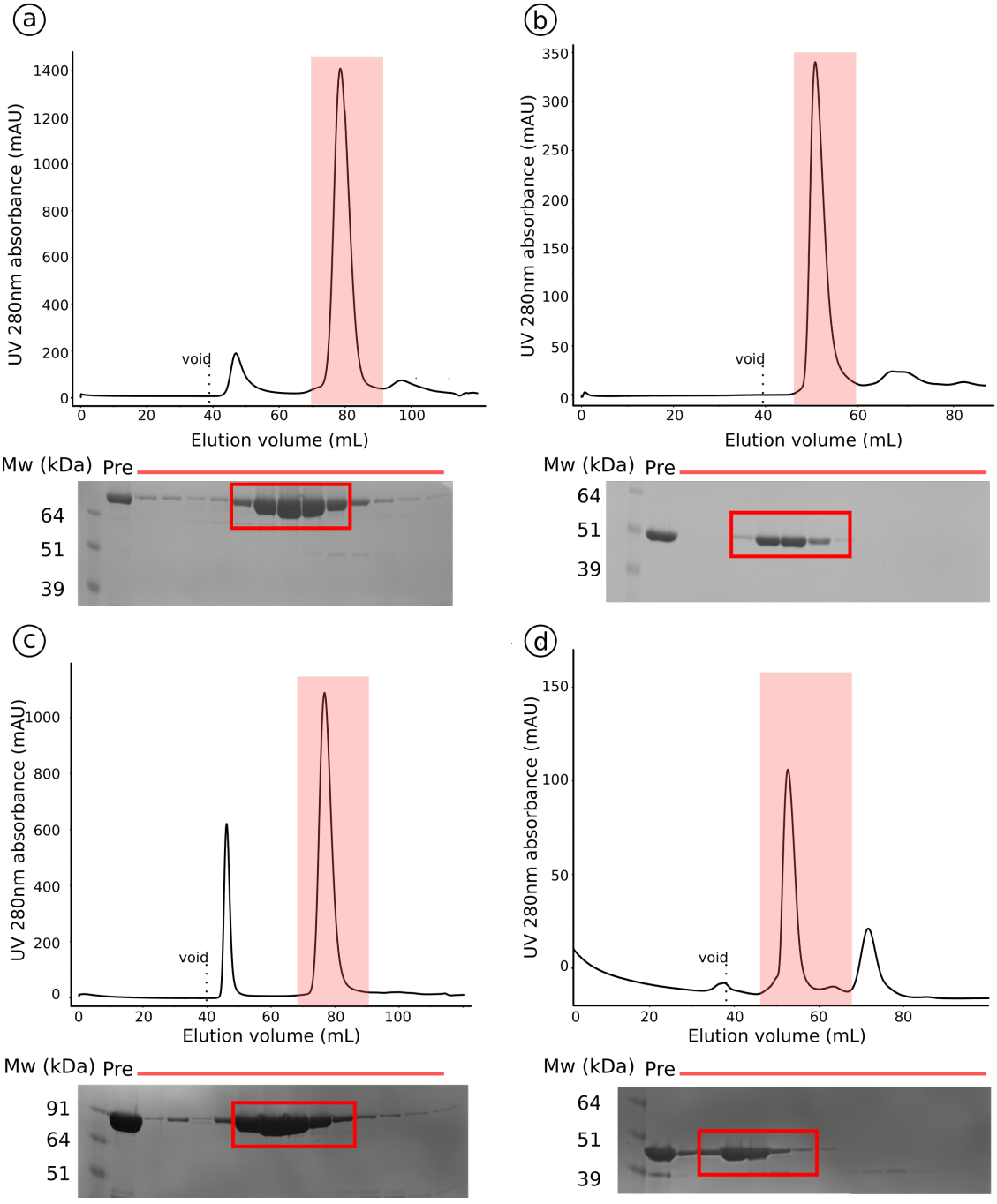



**Fig. 3.9 Purification of Treslin$_{\text{SHD-ext}}$ and Treslin$_{\text{ext}}$.** Chromatogram of final size-exclusion chromatography step and Coomassie-stained SDS PAGE of highlighted peak for **(a)** MBP-Treslin$_{\text{SHD-ext}}$, **(b)** Treslin$_{\text{SHD-ext}}$, **(c)** MBP-Treslin$_{\text{ext}}$, and **(d)** Treslin$_{\text{ext}}$. For (a) and (c), the column used was Superdex200 16/600, whereby for (b) and (d) was Superdex75 16/600. Red boxes indicate the fractions that were pooled for subsequent experiments. Pre = sample before size-exclusion chromatography.

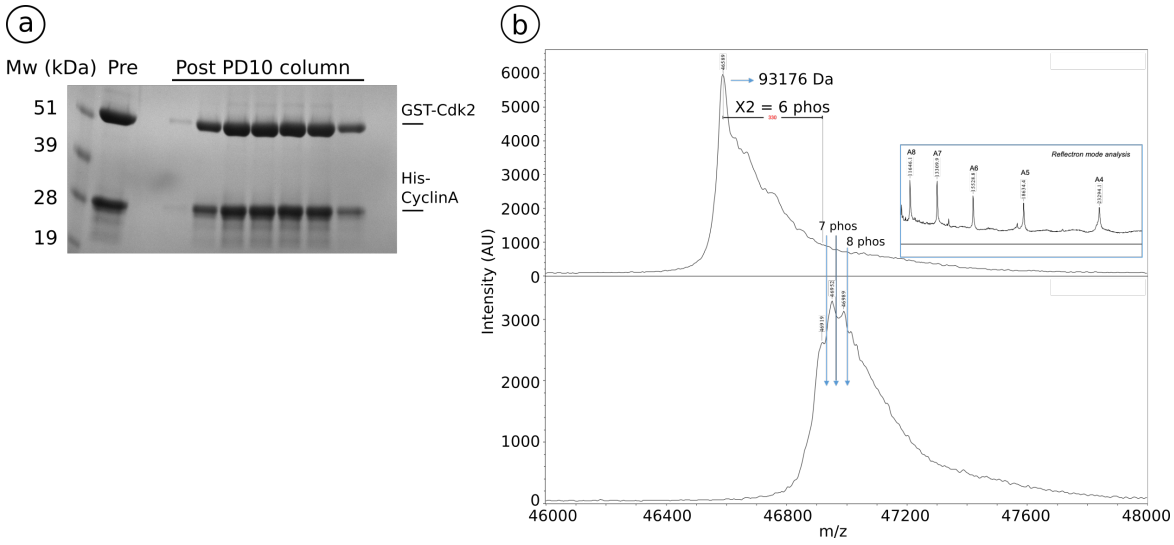



**Fig. 3.10 Purification of S-CDK and its activity on Treslin$_{SHD\text{-}ext}$.** **(a)** Coomassie-stained SDS PAGE of purified recombinant S-CDK. Pre = sample before PD10 column. **(b)** Result of MALDI-TOF MS for Treslin$_{SHD\text{-}ext}$ untreated (top panel) and treated with S-CDK (bottom panel). This figure was prepared by Dr Len Packman.

involved Ni-NTA affinity chromatography followed by Heparin-affinity chromatography in order to remove the DNA contaminants. When the N-terminal MBP tag was removed, the protein samples were incubated with TEV protease and subjected to another Ni-NTA affinity chromatography to capture the cleaved Treslin products. Both constructs, with or without MBP tag, were purified through a preparation grade size-exclusion chromatography column (figure 3.9). For each litre of culture, approximately 0.15 mg ($\pm$0.05 mg) of purified Treslin$_{SHD\text{-}ext}$ or Treslin$_{ext}$ could be obtained. This was in stark contrast with Treslin$_{SHD}$ expression and purification, which yielded 24 mg ($\pm$0.2 mg) of purified protein per litre of culture. Nevertheless, the purification process resulted in sufficient soluble and pure proteins, as judged by Coomassie-stained SDS-PAGE, either with or without N-terminal MBP tag for biochemical and biophysical experiments.

As mentioned previously, Treslin$_{ext}$ contains at least six potential CDK sites (section 3.2.5). To test whether the purified recombinant Treslin$_{ext}$ is a functional substrate for S-CDK, recombinant S-CDK was purified (figure 3.10b). The bacterial expression constructs were kind gifts from Dr Philip Zegerman of the Gurdon Institute, Cambridge. Subsequent MALDI-TOF MS analysis (performed by Dr Len Packman at the Department of Biochemistry, University of Cambridge) confirmed that Treslin$_{SHD\text{-}ext}$ is a functional substrate of S-CDK, as it was shown to be phosphorylated by S-CDK on at least six sites (figure 3.10b).

### 3.2.6 Treslin$_{\text{SHD-ext}}$ interaction with Cdc45

To test whether Treslin$_{\text{ext}}$ forms direct interaction with Cdc45, an MBP pulldown was performed. Purified MBP-tagged Treslin$_{\text{SHD-ext}}$ and Treslin$_{\text{ext}}$ were used as the 'bait' and bound to amylose resin (figure 3.11a). Purified Cdc45, in two times molar excess, was used as the target protein. The bound proteins were eluted from the beads by buffer containing 15 mM d-maltose. As shown in figure 3.11b, Cdc45 co-eluted from the beads with all constructs of Treslin. Cdc45 bound neither to the MBP tag nor to the amylose beads. Treslin$_{\text{SHD-ext}}$ was able to interact with Cdc45, as expected since the Treslin construct contains the putative Cdc45-binding domain in SHD. Surprisingly, the Treslin construct lacking the SHD domain (i.e. Treslin$_{\text{ext}}$), was also capable of binding to Cdc45 (figure 3.11b, lane 7). This result indicated that Treslin forms an additional contact with Cdc45 that is independent from the SHD. This was the first time that Treslin$_{\text{ext}}$ was shown to be involved in interaction with Cdc45.

Since Treslin$_{\text{ext}}$ is a substrate for S-CDK phosphorylation and contains a Cdc45-binding domain, it is possible that S-CDK might affect the interaction between Treslin$_{\text{ext}}$ and Cdc45. This hypothesis was tested by repeating the MBP pulldown experiment, but the MBP-tagged Treslin constructs were pre-phosphorylated by S-CDK prior to the addition of purified Cdc45.

The phosphorylated Treslin migrated slower compared to the non-phosphorylated proteins, as shown by the visible shift in the Coomassie-stained SDS-PAGE gel (figure 3.11c, lane 7 v 8, 9 v 10, and 11 v 12). The band shift in the gel can also be treated as evidence for phosphorylation of Treslin, as extra phosphates in the protein increases the negative charge of the protein — even in the presence of SDS — which altered the migration behaviour of the protein in the gel. There was no observable difference in amount of the Cdc45 co-eluting with MBP-tagged Treslin$_{\text{SHD}}$ or Treslin$_{\text{SHD-ext}}$ (figure 3.11c). However, there was a drastic change in Cdc45-binding as less Cdc45 co-eluted with phosphorylated MBP-tagged Treslin$_{\text{ext}}$ when compared to the non-phosphorylated MBP-tagged Treslin$_{\text{ext}}$ (figure 3.11c, lane 11 v 12). When Cdc45 and MBP-tagged Treslin complex were pre-formed prior to the S-CDK phosphorylation step, the result did not change (figure 3.11d). Taken together, these results demonstrated that S-CDK phosphorylation disrupts the Cdc45–Treslin$_{\text{ext}}$ interaction. S-CDK specifically targets the Treslin$_{\text{ext}}$ containing at least six potential S-CDK sites. This is an indication on how S-CDK may regulate Cdc45 and Treslin complex stability, by specifically targeting the Cdc45-binding region of Treslin$_{\text{ext}}$.

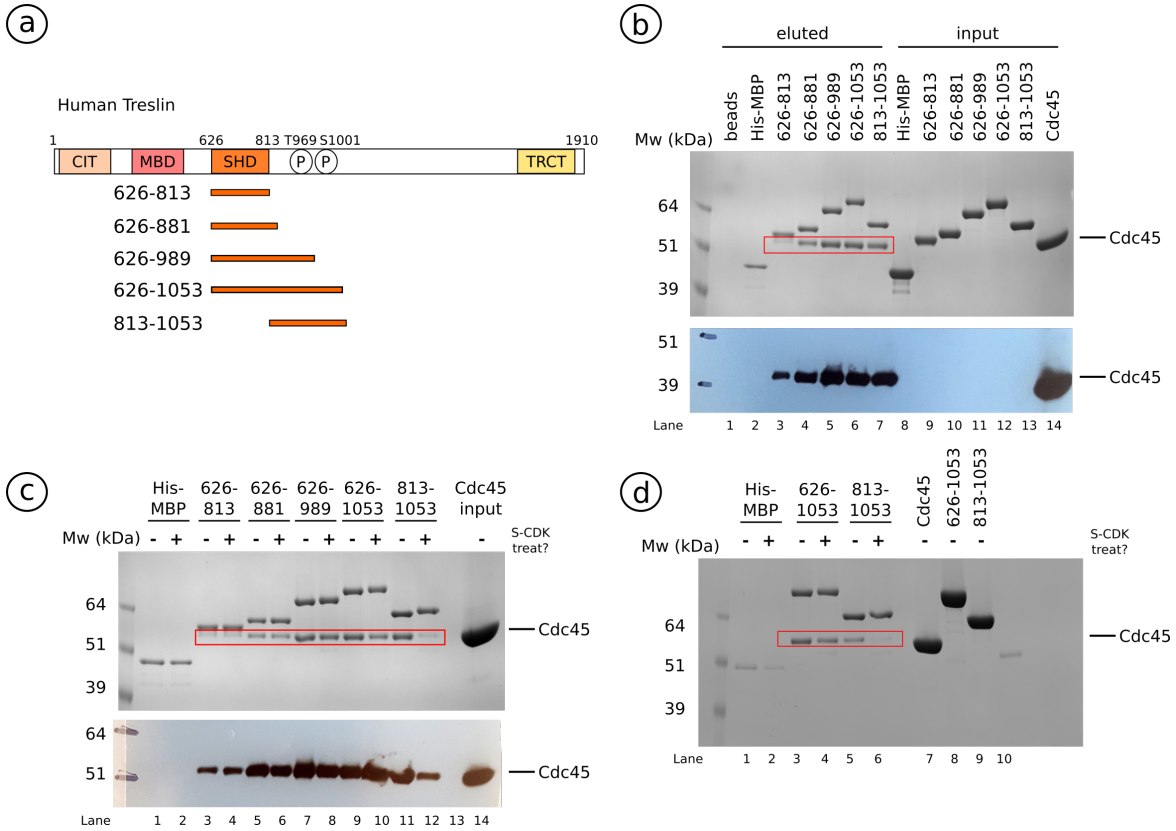



**Fig. 3.11 The effect of S-CDK phosphorylation on Treslin$_{\text{SHD-ext}}$ interaction with Cdc45.** **(a)** Schematic of MBP-tagged Treslin constructs used in these experiments. **(b)** Coomassie-stained SDS-PAGE of MBP pulldown using various MBP-tagged Treslin constructs with Cdc45. Co-eluted Cdc45 was highlighted in red box. Treslin residues 626–813 (lane 3) is poorly stained due to overlap with Cdc45. Bottom panel is the corresponding western blot using anti-Cdc45. **(c)** As in (b), but Treslin was treated with S-CDK before adding Cdc45. Bottom panel is the corresponding western blot using anti-Cdc45. **(d)** As in (c), except Cdc45 was added before S-CDK treatment and only Treslin residues 626–1053 and 813–1053 were used.

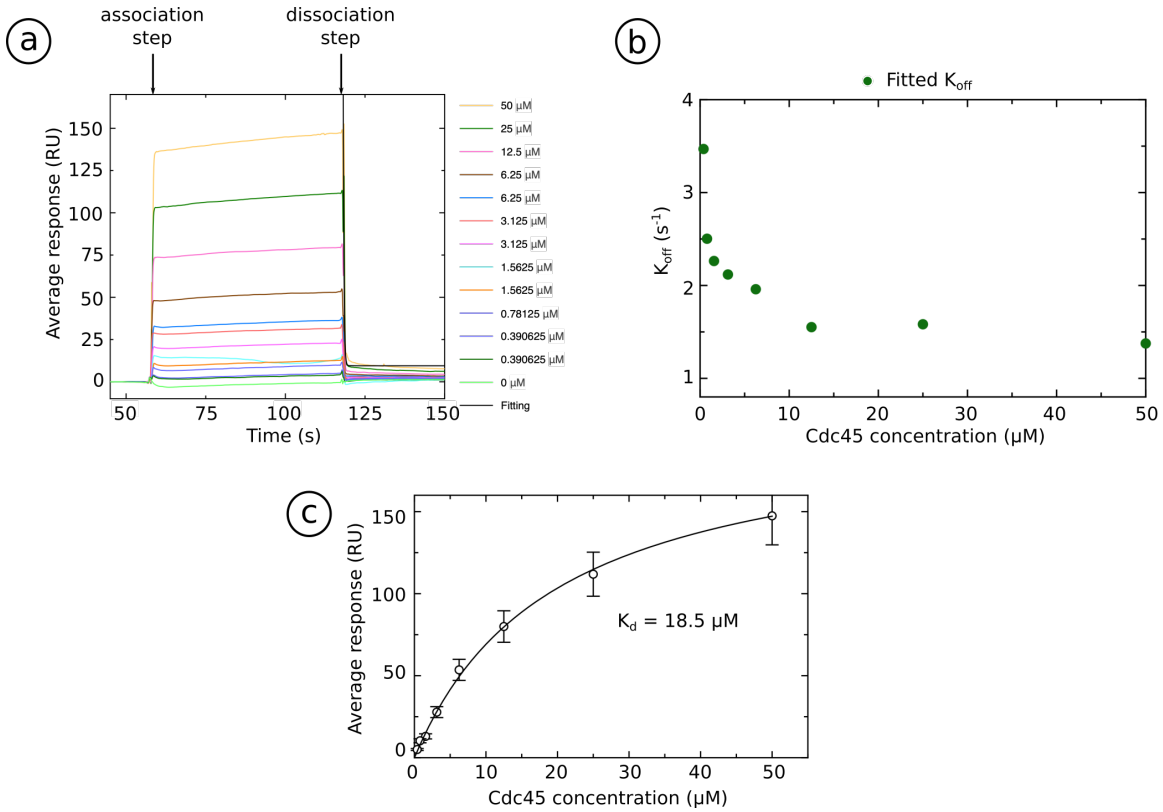



**Fig. 3.12 BLI measurements of Cdc45–Treslin$_{\text{SHD-ext}}$.** **(a)** Normalised sensorgram of Surface Plasmon Resonance (SPR) with first order decay fitting model. **(b)** Plot of fitted dissociation rate ($K_{off}$) against Cdc45 concentration. Average $K_{off}$ = 2.10 $s^{-1}$. **(c)** Steady state analysis using the results from the SPR measurement of Cdc45–Treslin$_{\text{SHD-ext}}$, based on two repeats.

### 3.2.7 Measurement of Cdc45–Treslin$_{SHD\text{-}ext}$ binding affinity and kinetics

SPR was performed using a Biacore T200 (GE Healthcare) system to measure the binding affinity and kinetics of the Cdc45–Treslin$_{SHD\text{-}ext}$ interaction. 15 nM of purified biotinylated MBP-tagged Treslin$_{SHD\text{-}ext}$ sample was immobilised on a gold-layered sensor chip containing streptavidin (XanTec BioAnalytics). The biotinylation of the MBP-tagged Treslin$_{SHD\text{-}ext}$ was confirmed by LC-MS and western blot using streptavidin-AP conjugated antibodies, as described previously (section 3.2.4).

Steady state analysis revealed that the mean $K_d$ of Cdc45–Treslin$_{SHD\text{-}ext}$ complex is 18.5 $\mu$M (figure 3.12a). This value is approximately two-fold higher compared to the $K_d$ of Cdc45–Treslin$_{SHD}$ complex — measured to be between 4.8 to 8.5 $\mu$M (figure 3.6). Intriguingly, the interaction of Cdc45–Treslin$_{SHD\text{-}ext}$ is characterised by very fast association and dissociation rates (figure 3.12b), reminiscent of the interaction observed between Cdc45 and Treslin$_{SHD}$ (figure 3.6). The dissociation rate ($K_{off}$) for each analyte concentration was fitted using a first order decay model. The values were then averaged ($K_{off}$ = 2.103 $s^{-1}$). It needs to be noted that there is no clear linear trend for the plot of $K_{off}$ against Cdc45 concentration (figure 3.12b). This value was then divided with the calculated $K_d$ to give the association rate ($K_{on}$) = 1.12 $\times$ $10^5$ $M^{-1}$ $s^{-1}$. These measurements showed that Treslin$_{ext}$ does not alter the kinetics of the Cdc45–Treslin complex, which suggested that Cdc45–Treslin interaction is characterised by has an inherently fast association and dissociation rates. Furthermore, the $K_d$ value suggested that the interaction of Cdc45–Treslin$_{SHD\text{-}ext}$ is two-fold weaker compared to the Cdc45–Treslin$_{SHD}$.

Similar results were also observed when performing the Bio-Layer Interferometry (BLI) measurement using the OCTET Red96 system (Pall ForteBio). The mean $K_d$ obtained from the steady state affinity analysis was 9.7 $\mu$M (figure 3.13b). Fitting of the dissociation step could not be performed as the analyte (Cdc45) did not dissociate from the immobilised ligand (figure 3.13c). The non-dissociation event was still observed even when changing the buffer condition, adding BSA and/or 0.005% (v/v) Tween 20, or long dissociation step (10, 30, and 60 mins. Data not shown). The steady state fitting in order to obtain the $K_d$ was only performed for the association step using a 1-to-1 model (figure 3.13a).

The results from the SPR and BLI measurements suggest that the binding affinity for the Cdc45–Treslin$_{SHD\text{-}ext}$ interaction (approximately 9.7 to 18.5 $\mu$M) is at least two-fold weaker compared to Cdc45–Treslin$_{SHD}$ (4.8 to 8.5 $\mu$M). This is intriguing, because two Cdc45-binding sites on Treslin should in theory make the binding tighter (i.e. lower $K_d$). It is possible that the binding in one site may weaken the conformation on another binding site, although this cannot be detected from these BLI and SPR measurements. The relatively

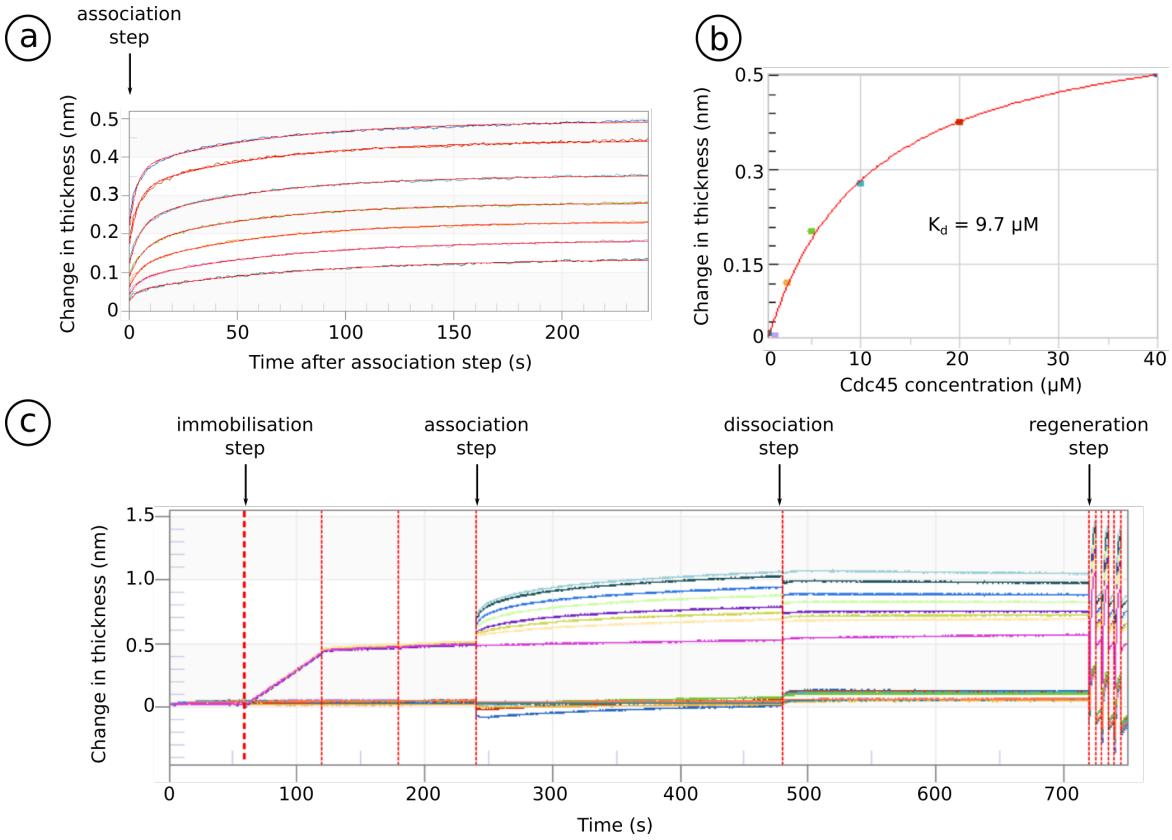



**Fig. 3.13 BLI measurement of Cdc45–Treslin$_{\text{SHD-ext}}$.** **(a)** Normalised sensorgram of Bio-Layer Interferometry (BLI) with partial fitting to the association step using 1-to-1 model. **(b)** Steady state analysis using results from the BLI measurement. **(c)** Raw BLI sensorgram for the experiment, indicating non-dissociation of the analyte (Cdc45). In this case, the dissociation step is 240 s.

weak $K_d$ of the complex may reflect the transient nature of the Cdc45–Treslin interaction in the cell.

The analyte (Cdc45) non-dissociation event observed during the BLI measurement (figure 3.13c) is also particularly intriguing. It is likely due to a reproducible experimental artefact caused by biosensor method of BLI. However, it may also be caused by yet an unknown biologically relevant phenomenon. Furthermore, attempts to repeat the SPR or BLI measurements using phosphorylated $Treslin_{SHD\text{-}ext}$ were unsuccessful due to very low yield after purification (data not shown).

### 3.2.8 $Treslin_{SHD\text{-}ext}$ is a monomer in solution

To determine the oligomerisation state of Treslin constructs containing the 'extension' domain, SEC-MALS was performed. Both $Treslin_{SHD\text{-}ext}$ and $Treslin_{ext}$ have very poor chromatographic behaviour in low salt buffer (160 mM NaCl), likely due to non-specific interactions with the column resin. This problem could be partly alleviated by increasing the salt concentration to 300 mM NaCl (figure 3.14a). Therefore, SEC-MALS for both $Treslin_{SHD\text{-}ext}$ and $Treslin_{ext}$ was performed in high salt condition (300 mM NaCl). It needs to be noted that for SEC-MALS of $Treslin_{SHD}$ (section 3.2.3) or any Treslin construct in complex with Cdc45 (section 3.2.9), low salt condition (160 mM NaCl) was used. S-CDK phosphorylation also improved the chromatographic behaviour of both $Treslin_{SHD\text{-}ext}$ and $Treslin_{ext}$ constructs. In addition, the $Treslin_{ext}$ sample had to be phosphorylated with S-CDK prior to the experiment as the non-phosphorylated $Treslin_{ext}$ sample failed to elute from the size-exclusion chromatography column — regardless of the salt concentration in the buffer — and therefore could not be analysed. In addition, MBP-tagged Treslin constructs were used for the experiments to increase the solubility of the sample, to improve the chromatographic behaviour, and to provide better resolution allowing better separation between the MBP-tagged Treslin and Cdc45 in the semi-analytical Superdex200 Increase 10/300 size-exclusion chromatography column (section 3.2.9).

MBP-tagged $Treslin_{SHD\text{-}ext}$ at 50 $\mu$M was applied to a Superdex200 Increase 10/300 size-exclusion chromatography column. It eluted as a single peak of 90.3 kDa ($\pm$0.1 kDa), close to its predicted monomeric size of 93.3 kDa (figure 3.14b). S-CDK-phosphorylated MBP-tagged $Treslin_{SHD\text{-}ext}$ also eluted as a single peak of 93.8 kDa ($\pm$0.43 kDa), indicating that phosphorylation by S-CDK does not alter the oligomeric state of $Treslin_{SHD\text{-}ext}$ (figure 3.14c). Meanwhile, phosphorylated MBP-tagged $Treslin_{ext}$ (at 50 $\mu$M) eluted as a single peak of 70.0 kDa ($\pm$0.07 kDa), close to its theoretical monomeric size of 71.9 kDa (figure 3.14d). Therefore, the SEC-MALS results demonstrated that Treslin constructs containing

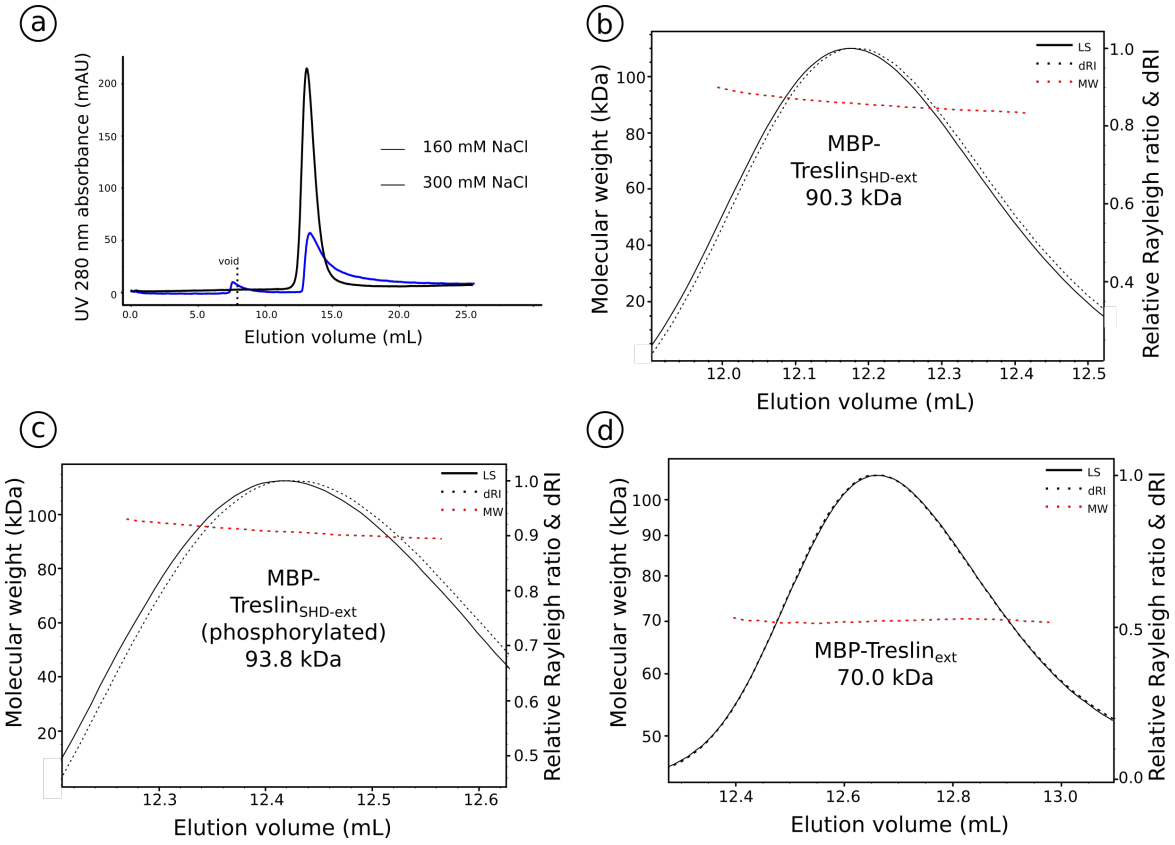



**Fig. 3.14 SEC-MALS of Treslin$_{SHD-ext}$ and Treslin$_{ext}$.** **(a)** A Superdex200 10/300 (GE Healthcare) size-exclusion chromatography of MBP-tagged Treslin$_{SHD-ext}$ in 160 (black line) or 300 mM NaCl (blue) concentration. **(b)** SEC-MALS of MBP-tagged Treslin$_{SHD-ext}$. Polydispersity = 1.000 ($\pm$0.269%). Expected Mw = 93.4 kDa. **(c)** SEC-MALS of S-CDK-phosphorylated MBP-tagged Treslin$_{SHD-ext}$. Polydispersity = 1.000 ($\pm$0.643%). Expected Mw = 93.4 kDa.**(d)** SEC-MALS of S-CDK-phosphorylated MBP-tagged Treslin$_{ext}$. Polydispersity = 1.000 ($\pm$0.139%). Expected Mw = 71.9 kDa. For (b), (c), and (d), the column used was Superdex200 Increase 10/300 (GE Healthcare). LS = Light scattering, Mw = Molecular weight, dRI = differential Refractometer index.

the 'extension' domain adopt a monomeric state in solution, regardless of the phosphorylation state.

### 3.2.9 Oligomeric state of the Cdc45–Treslin$_{\text{SHD-ext}}$ complex

The effect of S-CDK phosphorylation on the oligomeric state of Cdc45–Treslin$_{\text{SHD-ext}}$ complex was also tested by SEC-MALS. A 50 $\mu$M sample of Cdc45 was incubated with 50 $\mu$M of MBP-tagged Treslin$_{\text{SHD-ext}}$. In case of the phosphorylated sample, 50 $\mu$M of MBP-tagged Treslin$_{\text{SHD-ext}}$ was pre-incubated with 0.5 $\mu$M of S-CDK for 1 h at 25 °C before mixing it with Cdc45. The salt concentration in the buffer used for this experiment was 160 mM NaCl, since high salt condition inhibits Cdc45–Treslin complex formation (section 3.2.2).

Both samples eluted in two different peaks: one at 12 mL and a second at 14.2 mL (figure 3.15). The latter peak corresponded to monomeric Cdc45, with molecular weight of 65 to 66 kDa (predicted monomer size = 65.4 kDa). Intriguingly, the early peaks differed in the two samples. The first peak of the sample containing Cdc45 and phosphorylated MBP-tagged Treslin$_{\text{SHD-ext}}$ corresponds to a species with molecular weight of 136.0 kDa ($\pm$0.3 kDa) with polydispersity = 1.000. This result was 19.7 kDa short of a predicted 155.7 kDa value, assuming a 1-to-1 stoichiometry for Cdc45 and MBP-tagged Treslin$_{\text{SHD-ext}}$ (figure 3.15b). Meanwhile, the sample of Cdc45 and non-phosphorylated MBP-tagged Treslin$_{\text{SHD-ext}}$ produced an early peak with a rising molecular weight species. This could be explained by an aggregation over size-exclusion column (polydispersity = 1.016) and hence it did not form a stable complex (figure 3.15a) — although the elution point for this particular peak was later than the void volume (8 mL) of the column.

This observation was intriguing, considering that the Treslin$_{\text{ext}}$ interaction by itself is abrogated upon S-CDK phosphorylation (figure 3.11). It seems that Treslin$_{\text{SHD-ext}}$ forms a stable and stoichiometric complex during size-exclusion chromatography. However, it has to be noted that non-phoshorylated Treslin$_{\text{SHD-ext}}$ alone has a poor chromatographic behaviour on a Superdex200 10/300 size-exclusion chromatography column as it interacts non-specifically with the column's resin. This may be the main reason why the complex of non-phosphorylated Cdc45–Treslin$_{\text{SHD-ext}}$ was poorly resolved during SEC-MALS.

### 3.2.10 Treslin$_{\text{ext}}$ is an additional and weaker binding site of Cdc45

Results from the MBP pulldown experiment suggests that Treslin$_{\text{ext}}$ contains an additional and distinct Cdc45-binding site outside the SHD domain. S-CDK phosphorylation of Treslin$_{\text{ext}}$ was shown to weaken the Treslin$_{\text{ext}}$ interaction with Cdc45 (figure 3.11). However, it is not

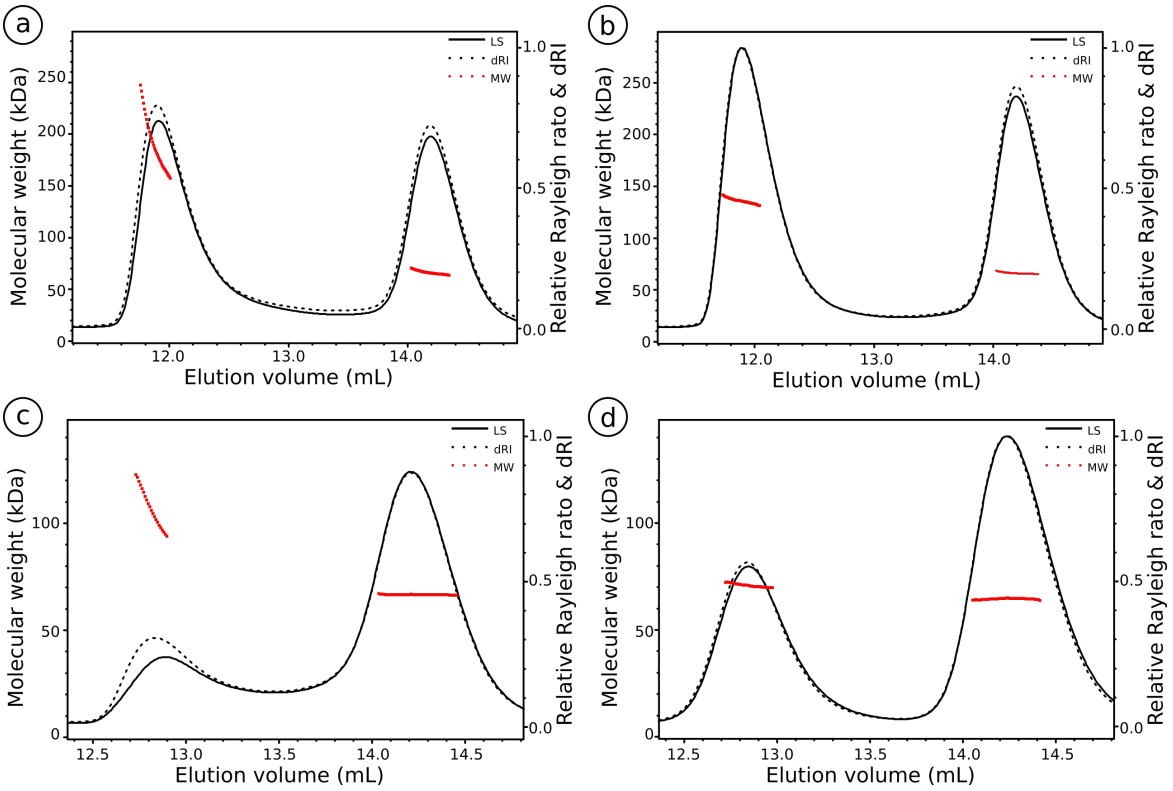



**Fig. 3.15 SEC-MALS of Cdc45–Treslin$_{\text{SHD-ext}}$ complex.** **(a)** SEC-MALS of Cdc45 and non-phosphorylated MBP-tagged Treslin$_{\text{SHD-ext}}$ complex. Polydispersity = 1.016 (±0.110%). Expected Mw = 155.7 kDa. **(b)** SEC-MALS of Cdc45 and S-CDK-phosphorylated MBP-tagged Treslin$_{\text{SHD-ext}}$ complex. Polydispersity = 1.000 (±0.312%). Expected Mw = 155.7 kDa. **(c)** SEC-MALS of Cdc45 and non-phosphorylated MBP-tagged Treslin$_{\text{ext}}$. Polydispersity = 1.007 (±0.214%). Expected Mw = 135 kDa. **(d)** SEC-MALS of Cdc45 and S-CDK-phosphorylated MBP-tagged Treslin$_{\text{ext}}$. Polydispersity = 1.000 (±0.118%). Expected Mw = 135 kDa. LS = Light scattering, Mw = Molecular weight, dRI = differential Refractometer index.

clear whether Treslin$_{ext}$ alone can sustain a complex formation with Cdc45. Two experiments were performed to test this possibility.

First, SEC-MALS described as in section 3.2.9 was repeated with an MBP-tagged Treslin$_{ext}$ construct. As expected, both samples eluted in two different peaks, early at 12.5 mL and late at 14.2 mL. The early peaks in the two samples differed in their molecular weight and polydispersity values. In both samples, the latter peak corresponded to excess monomeric Cdc45 with observed molecular weight of 64.6 to 66 kDa (predicted monomer size = 65.4 kDa).

If Cdc45 and Treslin$_{ext}$ co-eluted during the size-exclusion chromatography as a 1-to-1 stoichiometric complex, the size would be 135 kDa [65.0 kDa (Cdc45 monomer) + 70.0 kDa (MBP-tagged Treslin$_{ext}$ monomer)]. A mixture of Cdc45 and non-phosphorylated MBP-tagged Treslin$_{ext}$ produced an early peak corresponding to a species with a molecular weight size of 106.1 kDa ($\pm$1.2 kDa) with a polydispersity value of 1.007 (figure 3.15c). The poor polydispersity value could be explained by poor chromatographic behaviour of MBP-tagged Treslin$_{ext}$ by itself during a Superdex200 Increase 10/300 (GE Healthcare) size-exclusion chromatography in low salt buffer condition.

The early peak of the Cdc45 and phosphorylated MBP-tagged Treslin$_{ext}$ mixture showed a species with a molecular weight of 70.9 kDa ($\pm$0.16 kDa) and a polydispersity value of 1.000 (figure 3.15d). This observed molecular weight value was virtually identical with the expected 70.0 kDa of monomeric size of MBP-tagged Treslin$_{ext}$ by itself (figure 3.14d). Therefore, it could be concluded that the early peak was very likely to represent MBP-tagged Treslin$_{ext}$ alone (i.e. not in complex with Cdc45). This suggested that at this concentration, Treslin$_{ext}$ by itself is unable to co-fractionate with Cdc45 during the gel filtration, indicating weak protein-to-protein interactions. This is in contrast with Treslin$_{SHD}$ that is capable of co-eluting with Cdc45 during the gel filtration (figure 3.15b and d). Taken together, this indicates that Treslin$_{SHD}$ is the main interaction site with Cdc45, with the 'extension' domain acting as an additional but weaker binding site.

The binding affinity of Cdc45–Treslin$_{ext}$ was also measured by Bio-Layer Interferometry (BLI) using Octet RED96 system (Pall ForteBio). As with previous measurement using Treslin$_{SHD\text{-}ext}$ construct, the dissociation step of the Cdc45 (analyte) could not be observed (figure 3.16c). The $K_d$ was derived from the steady state affinity — using a 1-to-1 fitting of the association step — that gave a value of 14.8 $\mu$M (figure 3.16a and b). This measurement is consistent with the results from the SEC-MALS experiment, in which the Treslin$_{ext}$ containing a weaker Cdc45-binding site when compared to Treslin$_{SHD}$. However, the reproducible non-dissociation is intriguing as it might be caused by yet unknown biochemical phenomenon and/or experimental artefact (section 3.2.7). An SPR measurement was attempted to verify

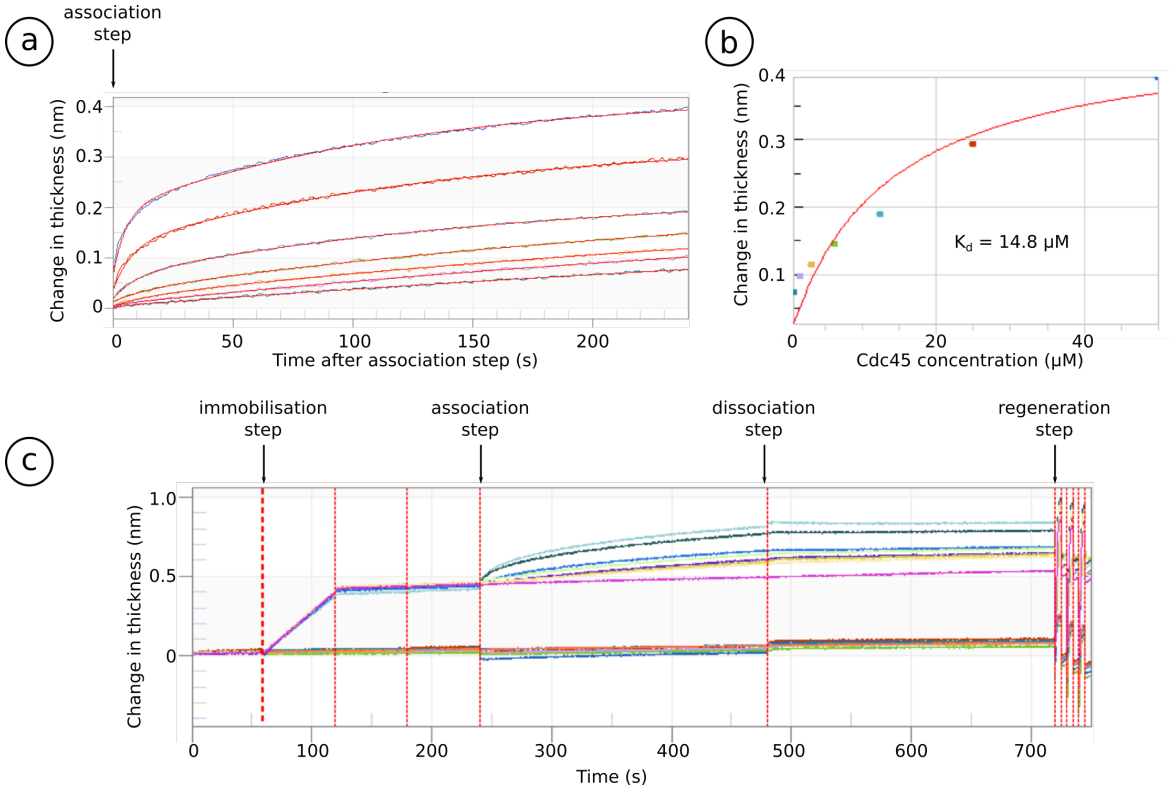



**Fig. 3.16 BLI analysis of Cdc45 and Treslin$_{ext}$ complex.** **(a)** Normalised sensorgram of Bio-Layer Interferometry (BLI) analysis with a 1:1 partial fitting model to the association step. **(b)** Steady state analysis using the results from the BLI experiments. **(c)** Raw sensorgram of the BLI in this experiment, showing the non-dissociation state of the analyte (Cdc45).

the result, however the resonance signal was not satisfactory due to non-specific interaction with the sensor chip regardless of buffer condition and having 0.005% (v/v) Tween 20 and/or BSA in the buffer (data not shown)

### 3.2.11 HDX–MS analyses of Cdc45–Treslin complex

A Hydrogen Deuterium Exchange coupled to Mass Spectrometry (HDX–MS) experiment was performed to map the region of Cdc45 that interacts with Treslin. This experiment was performed in collaboration with Dr Mark Skehel and Dr Sarah Maslen at the Laboratory of Molecular Biology, Cambridge. The basis of HDX–MS is that the rate of exchange from hydrogen in the amides of the protein backbone to deuterium — an isotope of hydrogen that contains a neutron — gives an information on which region is exposed to the aqueous solution. Each sample of Cdc45 alone, Cdc45–Treslin$_{SHD}$, Cdc45–Treslin$_{SHD\text{-}ext}$, and Cdc45–phosphorylated Treslin$_{SHD\text{-}ext}$ was incubated with buffer containing deuterium, digested with protease, and analysed by mass-spectrometry. To locate which region of Cdc45 binds to Treslin, the protected peptide sequences of Cdc45 in the absence or presence of Treslin were compared.

Cdc45 residues 419 to 434 were more protected from deuterium exchange when in complex with Treslin$_{SHD}$ (figure 3.17a). Based on the available structure of human Cdc45 (PDB: 5DGO), residues 419–434 fold into an $\alpha$-helix structure ($\alpha$-17, based on the structure's annotation) that contains three highly-conserved positively-charged amino acid residues (Simon et al., 2016). These residues are lysine 420 (K420), lysine 421 (K421), and arginine 424 (R424) (figure 3.17b).

To test whether these three positively-charged residues are essential for Treslin$_{SHD}$ binding, each one of them were mutated to a non-polar alanine. Interestingly, when K420 was mutated to alanine (K420A), Cdc45 was unable to be expressed well in bacteria (figure 3.17c). It was likely that the Cdc45 K420A mutant was unable to fold correctly, suggesting an essential role of K420 for Cdc45's overall architecture. The other single mutants including K421A and R424A were expressed and able to be purified as well as the wild-type (data not shown). When tested in a GST pulldown experiment, both K421A and R424 mutants were able to interact with His-GST tagged Treslin$_{SHD}$ equally well as the wild-type Cdc45 (figure 3.17d). Although it is not clearly determined whether any of these residues are solely responsible for Treslin-binding, the Treslin-binding region of Cdc45 was likely to be mapped. It is likely that a combination of K420, K421 and R424 residues are functionally equivalent for Treslin$_{SHD}$ binding.

The observation that only $\alpha$-17 region (residue 419–434) of Cdc45 forms an interaction with Treslin may explain why the Cdc45–Treslin$_{SHD}$ complex has very fast kinetics. The

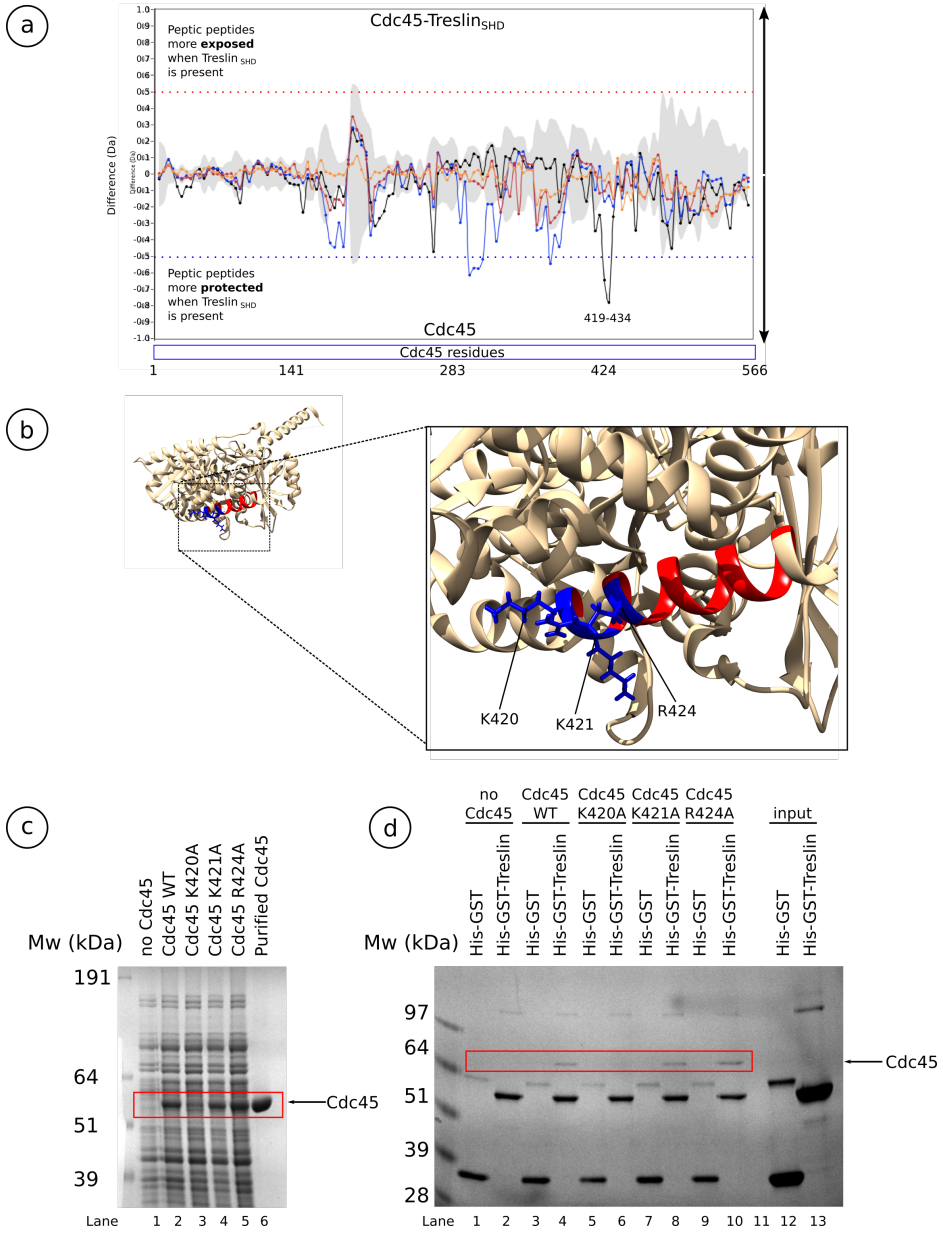



**Fig. 3.17 HDX-MS (Hydrogen-Deuterium Exchange–Mass Spectrometry) of the Cdc45–Treslin$_{SHD}$ complex.** **(a)** Data difference plot Cdc45 vs Cdc45–Treslin$_{SHD}$. Grey = error range, Orange = 3 s, Red = 30 s, Blue = 300 s, Black = 3000 s. Figure was made by Dr Mark Skehel and Dr Sarah Maslen. **(b)** Region of Cdc45 ($\alpha$-helix 17th, based on the annotated structure) that was protected by Treslin$_{SHD}$ was indicated in red and blue. The blue highlighted the positively-charged amino acid residues in the helix. The side-chain of each residue is displayed. Inset = structure of Cdc45. PDB: 5DGO (Simon et al., 2016). **(c)** Coomassie-stained SDS-PAGE showing soluble lysate of Cdc45 and its alanine mutants. **(d)** Coomassie-stained SDS-PAGE showing GST pulldown of purified GST-tagged Treslin$_{SHD}$ and the soluble lysates of Cdc45 or its alanine mutants.

very fast on- and off-rate observed in the separate BLI and SPR measurements (section 3.2.4 and figure 3.6) may be due to the fact that this interaction involves only a short peptide to peptide contact.

Cdc45 was being more protected by Treslin$_{\text{SHD-ext}}$ during the HDX-MS experiment, confirming that Treslin$_{\text{SHD-ext}}$ binds more extensively to Cdc45. The data difference plot showed that residues 419–434 (also protected by Treslin$_{\text{SHD}}$ alone), 188–199, 402–434, 467–476, 500–525, and 547–560 were protected by Treslin$_{\text{SHD-ext}}$ (figure 3.18). S-CDK phosphorylation on Treslin$_{\text{SHD-ext}}$ did not produce any noticeable difference in the Cdc45's protection (figure 3.18a and b). This might be caused by the inability of the technique to resolve the difference, especially when the peptide coverages for Cdc45–Treslin$_{\text{SHD-ext}}$ (phosphorylated) were less than Cdc45–Treslin$_{\text{SHD-ext}}$ (non-phosphorylated). Nevertheless, this result showed that Treslin$_{\text{SHD-ext}}$ binds to Cdc45 more extensively compared to Treslin$_{\text{SHD}}$ alone.

### 3.2.12 Structural analyses of the Cdc45–Treslin complex

X-ray crystallography experiments were initially performed to obtain the structure of Cdc45–Treslin complex. First, a Cdc45 and Treslin$_{\text{SHD}}$ (residues 626–813) complex at 400 $\mu$M was reconstituted either by purifying it over size-exclusion chromatography column or simply mixing it. Crystal trays were set up by using five commercial screens: JCSG+, Morpheus, PACT, ProPlex, and Wizard I & II (Molecular Dimensions) at various temperatures (either at 4, 11, or 19 °C). Either wild-type or a crystallisation construct (Δ154–164) of Cdc45 was used for the experiment. After incubating for more than eight weeks, no protein crystal was observed in any of the conditions.

The experiments were repeated using Treslin$_{\text{SHD-ext}}$. The complex of Cdc45–Treslin$_{\text{SHD-ext}}$ at 100 $\mu$M (either with wild-type or crystallisation construct of Cdc45) was either purified by size-exclusion chromatography or simply mixed prior to setting up crystal trays. The crystal trays conditions were the same as described in previous paragraph. Again, no protein crystal was formed. These unsuccessful attempts to crystallise the complex could be explained by the complex's dynamic nature and rapid kinetics that could prevent stable crystal packing (figures 3.6 and 3.12). The limited amount of purified Treslin$_{\text{SHD-ext}}$ also made it difficult to obtain more concentrated samples that might be required to form a stable crystal of the complex.

Single-particle cryo-electron microscopy (cryo-EM) has become a mainstream choice for obtaining high-resolution protein structures. This is driven by advancements in various aspects of cryo-EM technologies, such as the development of a more sensitive direct electron detection camera with higher detective quantum efficiency (DQE), camera's sensors that

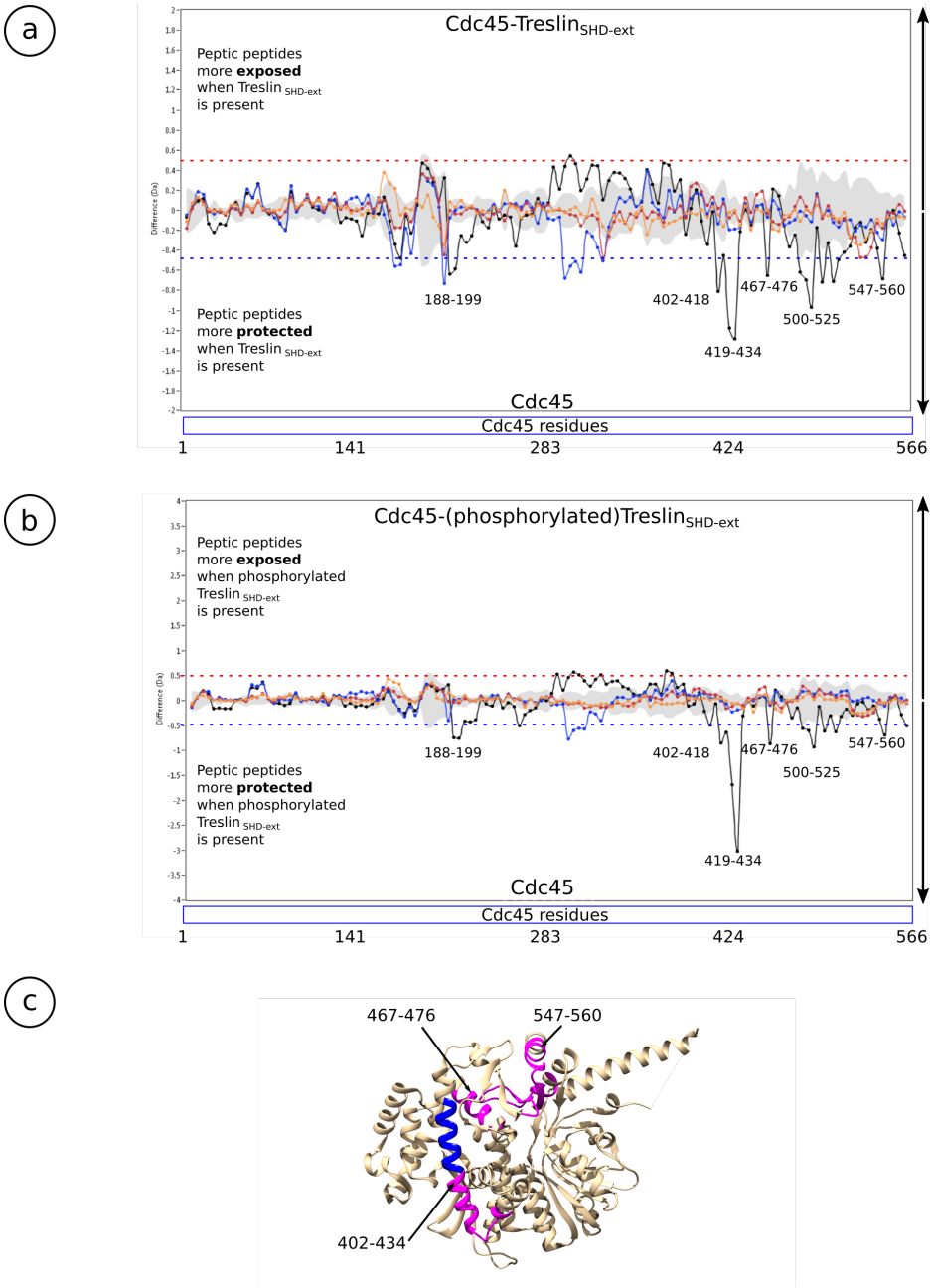



**Fig. 3.18 HDX-MS (Hydrogen-Deuterium Exchange–Mass Spectrometry) of the Cdc45–Treslin$_{SHD-ext}$ complex. (a)** Data difference plot Cdc45 vs Cdc45–Treslin$_{SHD-ext}$. **(b)** Data difference plot Cdc45 vs Cdc45–Treslin$_{SHD-ext}$ (phosphorylated). Grey = error range, Orange = 3 s, Red = 30 s, Blue = 300 s, Black = 3000 s. **(c)** Region of Cdc45 protected by Treslin$_{SHD-ext}$. Blue = protected by both Treslin$_{SHD}$ and Treslin$_{SHD-ext}$. Magenta = protected by Treslin$_{SHD-ext}$ only. PDB: 5DGO (Simon et al., 2016). Figures (a) and (b) were prepared by Dr Sarah Maslen and Dr Mark Skehel.

are able to record multiple-frame movies, improvements in data processing algorithm, and more powerful computing power (Cheng, 2018; McMullan et al., 2016; Nogales and Scheres, 2015). Previously, single-particle cryo-EM analysis was limited to large protein samples (approximately >150 kDa) but recent development of the Volta Phase Plate (VPP) expanded the possibility of analysing smaller proteins. VPP introduces additional low spatial frequency signal which improves the contrast and hence improves slightly particle picking (Danev and Baumeister, 2017). Furthermore, the inherent advantages of single-particle cryo-EM analysis include bypassing the requirement of large amount of protein samples and obtaining well-diffracted protein crystals. Therefore, the focus to obtain structure of the Cdc45–Treslin complex was redirected to single-particle cryo-EM analysis.

### Cryo-EM analysis of the Cdc45–Treslin$_{\text{SHD-ext}}$ complex

Purified Cdc45 and Treslin$_{\text{SHD-ext}}$ were mixed and applied to a semi-analytical Superose6 Increase 3.2/300 size-exclusion chromatography column (GE Healthcare). Treslin$_{\text{SHD-ext}}$ sample was pre-phosphorylated with S-CDK before incubating with Cdc45 to improve the chromatographic behaviour. The complex eluted from the column and appeared stoichiometric judged by Coomassie-stained SDS-PAGE (figure 3.19a).

Superose6 Increase 3.2/300 may not be the best column to separate the complex from the individual protein components, as it did not have baseline separation for both protein species (figure 3.19a). However, there were several reasons for using this column. First, the small volume of the column helped to avoid a large dilution effect, which ensures that the complex concentration is well above the $K_d$ throughout the size-exclusion chromatography run. Secondly, Treslin$_{\text{SHD-ext}}$ has lesser tendency to interact non-specifically to the Superose resin than to Superdex resin (data not shown). Lastly, this size-exclusion chromatography served as a buffer exchange step to ensure uniformity in the buffer for cryo-EM sample preparation.

The protein complex sample at 5 $\mu$M was applied to a glow discharged R 0.6/1.0 UltrAuFoil 300-mesh holey gold grid (Quantifoil). The grid was used for data acquisition using a Titan Krios microscope (FEI ThermoFisher) operated at 300 kV, equipped with a Volta Phase Plate and K2 Summit direct electron detector (Gatan Inc.) in 'counting mode'.

132,610 particles were automatically picked from 1,170 motion-corrected micrographs (figure 3.19b). After two rounds of reference-free 2D classifications, four satisfactory class averages (total 24,827 particles) showed secondary structure features (figure 3.19c). The corresponding particles were kept for 3D classification. After generating the initial model followed by a 3D classification, the subsequent particles only produced a 3D reconstructed map at 10 Å resolution (figure 3.19e). There was no case of preferred orientation or randomly

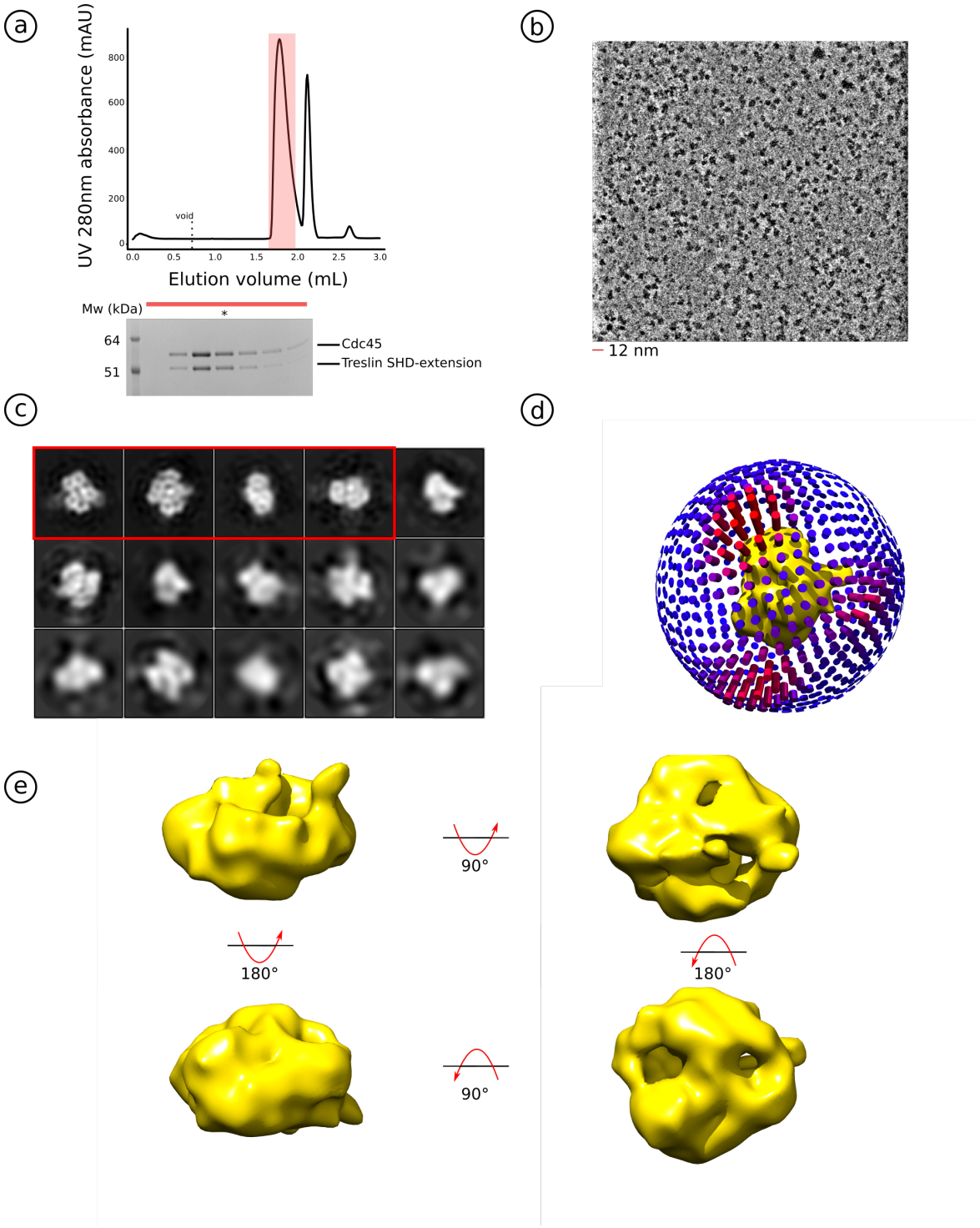



**Fig. 3.19 Cryo-EM analysis of the Cdc45–Treslin$_{\text{SHD-ext}}$ complex.** **(a)** Chromatogram of Superose6 Increase 3.2/300 and Coomassie-stained SDS-PAGE showing Cdc45–Treslin$_{\text{SHD-ext}}$ complex for sample preparation. The fraction indicated with asterisk was used for grid preparation. **(b)** Representative micrograph of the Cdc45–Treslin$_{\text{SHD-ext}}$ specimen, lowpass-filtered to 15 Å and highpass-filtered to 200 Å . **(c)** Top 15 class averages, ranked by the abundance of particles, after three rounds of 2D classification. Top 4 averages (highlighted in red box) were used for subsequent steps. **(d)** Angular distribution of the 3D map. Red and blue represent high and low number of particles contributing to the map, respectively. **(e)** 3D density map of Cdc45–Treslin$_{\text{SHD-ext}}$.

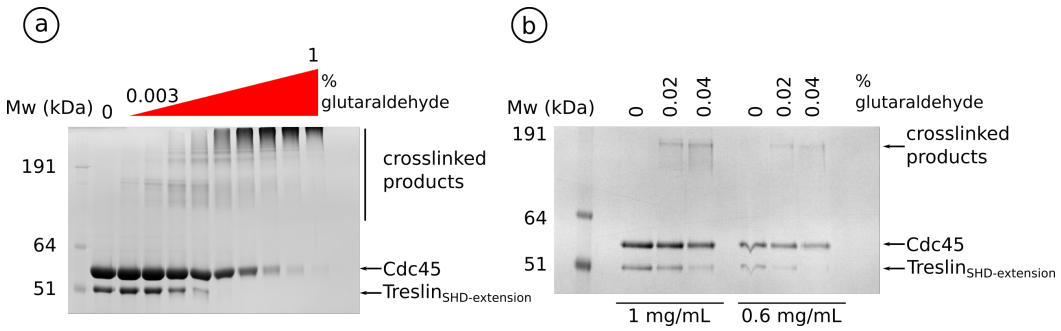



**Fig. 3.20 Glutaraldehyde crosslinking of the Cdc45–Treslin$_{SHD\text{-}ext}$ complex.** **(a)** Coomassie-stained SDS-PAGE showing titration of glutaraldehyde to crosslink Cdc45-Treslin$_{SHD\text{-}ext}$ complex. **(b)** Coomassie-stained SDS-PAGE gel showing cross-linking products of Cdc45 and Treslin$_{SHD\text{-}ext}$ at 1 or 0.6 mg/mL with different concentrations of glutaraldehyde.

distributed angular distribution (figure 3.19d). The poorly resolved 3D map was likely to be caused by a combination of sample heterogeneity (mixture of the protein complex and individual protein components) and inability of the RELION 3.0 software package to align the particles accurately. With such low resolution map, it is difficult to unambiguously assign the location of Cdc45 and Treslin.

Sample heterogeneity could be caused by two main reasons. Firstly, as described earlier in section 3.2.7, the complex of Cdc45–Treslin$_{SHD\text{-}ext}$ has very fast kinetics. The observation of a seemingly stoichiometric Cdc45–Treslin$_{SHD\text{-}ext}$ complex after size-exclusion chromatography was possibly due to these two proteins migrating at almost similar rate in the column, which allowed them plenty of chances to repeatedly dissociate and re-associate. Unfortunately, the Superose6 Increase 3.2/300 column used did not have the capacity to fully resolve the complex and individual proteins.

Secondly, the heterogeneity could be due to protein complex disruption by the air-water interface during the grid freezing step (Noble et al., 2018). To resolve this issue, a number of attempts were made. Firstly, protein cross-linking was performed by using glutaraldehyde. Unfortunately, most of the protein complexes were not crosslinked as a single uniform band as indicated by Coomassie-stained SDS-PAGE (figure 3.20). Next, small concentration of detergent in the buffer (0.005% (v/v) Tween 20 in the buffer) and/or glutaraldehyde [0.005 or 0.01% (v/v)] were applied before the grid vitrification step to reduce preferred orientation issue. Lastly, the sample was also applied to a 300-mesh holey grids with a thin layer of graphene oxide (Palovcak et al., 2018). None of these attempts, either individually or in combination, produced a more well-behaved sample for single-particle cryo-EM analysis.

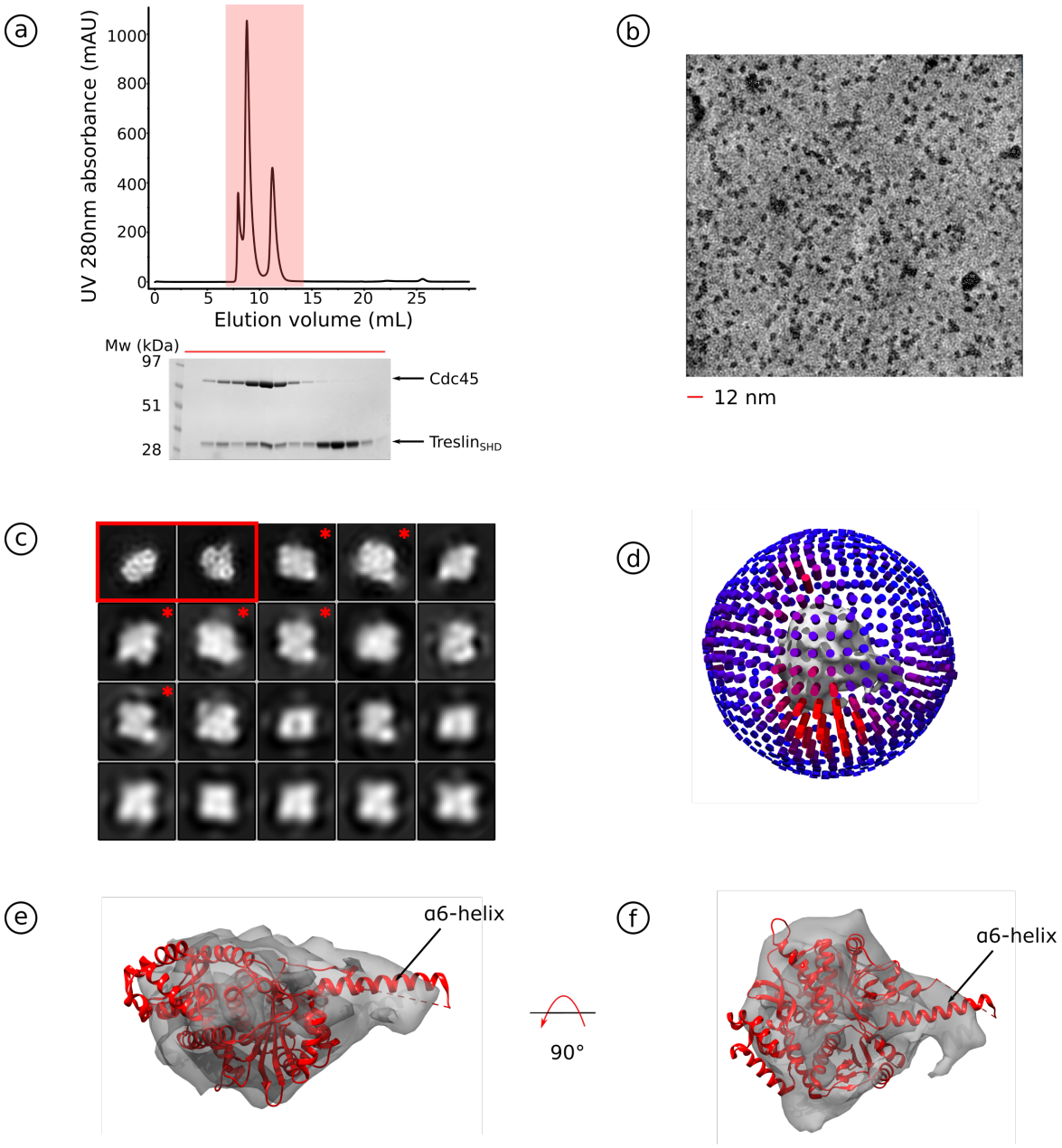



**Fig. 3.21 Cryo-EM analysis of the Cdc45–Treslin$_{SHD}$ complex. (a)** Chromatogram of Superdex75 10/300 size-exclusion chromatography run and Coomassie-stained SDS-PAGE showing Cdc45–Treslin$_{SHD}$ complex sample preparation. **(b)** Representative micrograph of the Cdc45–Treslin$_{SHD}$ specimen. **(c)** Top 20 class averages, ranked by the abundance of particles, after three rounds of 2D classification. Particles in the top two class averages (highlighted in red) were used for subsequent 3D reconstruction. Red asterisk indicates possible low-resolution alignment of Cdc45–Treslin$_{SHD}$ complex. **(d)** Angular distribution of the resulting 3D map. Red and blue represent high and low number of particles contributing to the map, respectively. **(e–f)** 3D density map reconstructed from particles saved in (c). **(g)** Fitting the crystal structure of human Cdc45 (highlighted in red, PDB: 5DGO) into the 3D density map (Simon et al., 2016).

### Cryo-EM analysis of the Cdc45–Treslin$_{SHD}$ complex

Purified Cdc45–Treslin$_{SHD}$ sample was applied to a semi-analytical Superdex75 10/300 size-exclusion chromatography column. The main peak fraction appeared to be a stoichiometric complex (figure 3.21a). The main peak sample at 2.2 $\mu$M was subsequently applied to a glow discharged UltrAufoil R0.6/1.0 holey gold grid (Quantifoil). After blotting the excess liquid, the grid was rapidly vitrified in liquid ethane. The resulting grid was used for data acquisition using a Titan Krios electron microscope (FEI ThermoFisher) operated at 300 kV and equipped with a Volta Phase Plate and a K2 Summit direct electron detector (Gatan Inc.) running in 'counting mode'.

393,780 particles from 1,521 motion-corrected micrographs (figure 3.21b) were automatically picked and extracted using the RELION 3.0 software package (Scheres, 2016). After two rounds of 2D classification, two distinct class averages of the sample could be observed (figure 3.21c, highlighted in red box). These averages, comprising 78,788 particles or 90% of total particles, showed distinct secondary structure features. However, after measuring the dimension of the averaged class by the pixel values (approximately 40 $\times$ 50 Å), it was clear that these averages belonged to Cdc45 alone. When these particles were used for subsequent 3D classification step, it produced a low-resolution 3D reconstruction map of Cdc45 (figure 3.21e–f). This might be caused by sample heterogeneity, preferred orientation (figure 3.21d), and inability of the RELION 3.0 software package to align the particles reliably. Docking of the Cdc45's crystal structure (PDB: 5DGO) to the density map showed no extra density that could be accounted for Treslin$_{SHD}$. The rest of the particles (figure 3.21c, highlighted by red asterisks) could belong to the complex of Cdc45–Treslin$_{SHD}$ that remained after vitrification step. Nevertheless, the 2D averages of these particles were poorly-defined and could not be used to reliably perform further processing.

To verify and improve the Cdc45's 3D reconstructed map obtained from RELION 3.0 software, the same dataset was re-processed using the cryoSPARC 2.8.0 software package (Punjani et al., 2017). 393,210 particles from 1,521 motion-corrected micrographs were automatically picked and extracted using the cryoSPARC software package. After two rounds of 2D classification, 71,293 particles were retained for ab-initio model generation and homogenous refinement steps (figure 3.22a). The 3D reconstructed map at 8 Å resolution obtained from cryoSPARC was slightly sharper and more interpretable than the one obtained in RELION (figure 3.22b-d). The crystal structure of Cdc45 (PDB: 5DGO) could be fitted well to the density (figure 3.22e) (Simon et al., 2016). In summary, this Cdc45–Treslin$_{SHD}$ dataset failed to produce a well-behaved sample as most of the complex fell apart during vitrification process. Nevertheless, a low-resolution 3D reconstruction map of Cdc45 could be obtained from this dataset (figures 3.21 and 3.22).

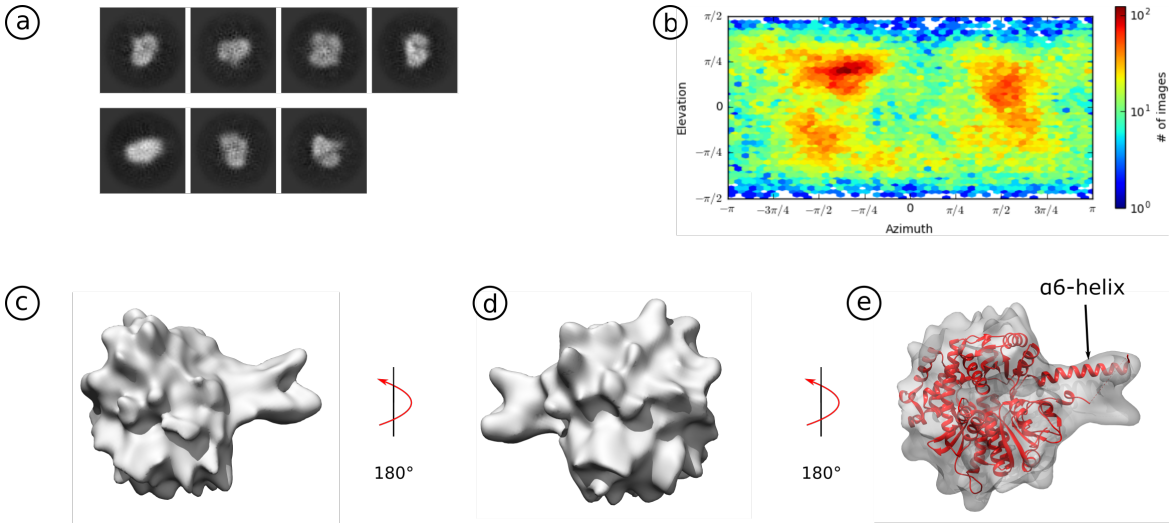



**Fig. 3.22 Refinement of Cdc45 structure using cryoSPARC software package.** **(a)** 2D class averages of the particles used in the 3D reconstruction **(b)** Angular distribution of the resulting 3D map. **(c–d)** 3D reconstructed map of Cdc45. **(e)** Fitting the crystal structure of human Cdc45 (highlighted in red, PDB: 5DGO) into the 3D density map (Punjani et al., 2017; Simon et al., 2016).

### Cryo-EM analysis of the Cdc45–Treslin$_{SHD}$ complex with glutaraldehyde

A small concentration of cross-linker, such as glutaraldehyde, may help to prevent complex disruption and/or preferred orientation caused by air-water interface during grid vitrification (Goswami et al., 2018). The previous experiment was repeated by applying 0.005% (v/v) glutaraldehyde to the purified complex before applying to the grid. Both protein purification and image acquisition steps were identical to those described earlier (figure 3.21a).

398,556 particles from 1,592 motion-corrected micrographs were automatically picked and extracted using the RELION 3.0 software package (figure 3.23a). After 2D classification, two different protein species could be observed. Majority of the class averages (58,273 particles) belonged to the 'small' particles (approximately 40 × 50 Å), likely to be Cdc45 by itself (figure 3.23b). The remaining averages (25,556 particles) grouped as the minority or the 'big' particles (approximately 70 × 100 Å), which were likely to be the Cdc45–Treslin$_{SHD}$ complex (figure 3.23c). The 'big' particles were subjected to a further round of 2D classification, retaining only 25,556 particles for further processing.

The resulting low-resolution 3D reconstructed map at approximately 12 Å showed a large density that could accommodate both Cdc45 and Treslin$_{SHD}$ (figure 3.23d–f). This was the first evidence that crosslinking experiment worked to a certain extent, as it helped to sustain complex formation after the vitrification step. However, in a such low-resolution map, it is not possible to unambiguously assign the position of Cdc45 and Treslin$_{SHD}$. This could be caused by a small number of good quality particles which made it very difficult for the

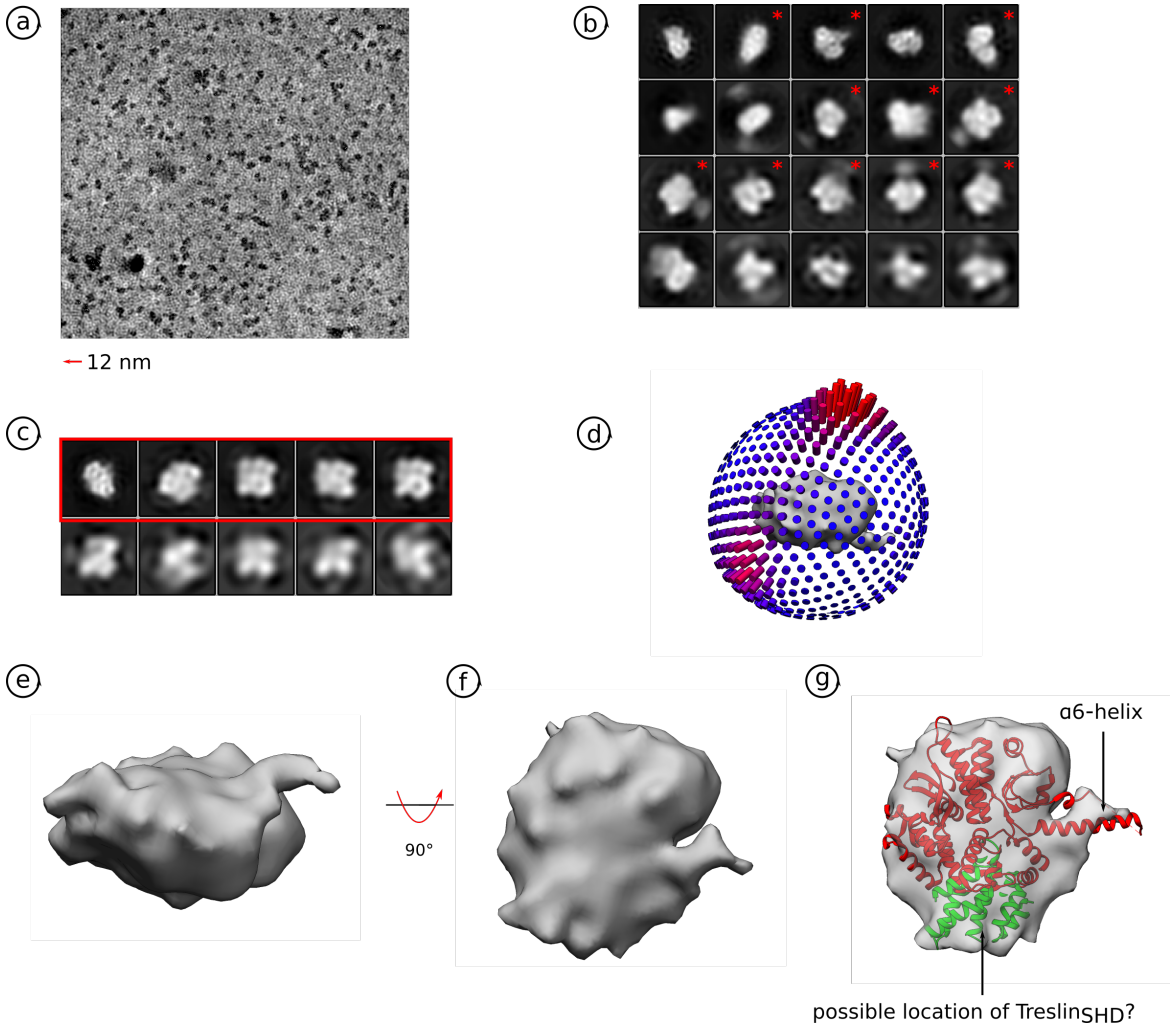



**Fig. 3.23 Cryo-EM analysis of the Cdc45–Treslin$_{SHD}$ complex with 0.005% (v/v) glutaraldehyde. (a)** Representative micrograph of the Cdc45–Treslin$_{SHD}$ specimen with 0.005% (v/v) glutaraldehyde. **(b)** Top 20 class averages, ranked by the abundance of particles, after three rounds of 2D classification. Averages highlighted with red asterisks referred to 'big' particles (Cdc45–Treslin$_{SHD}$ complex) and used for another round of 2D classification. **(c)** 2D classification result of 'big' particles. The particles in the averages highlighted in red box were used for subsequent 3D classification step. **(d)** Angular distribution of the resulting 3D map. Red and blue represent high and low number of particles contributing to the map, respectively. **(e–f)** 3D density map reconstructed from particles saved in (c). **(g)** Fitting the crystal structure of human Cdc45 (highlighted in red, PDB: 5DGO) and Treslin$_{SHD}$ (highlighted in green, unpublished data) into the 3D density map (Simon et al., 2016).

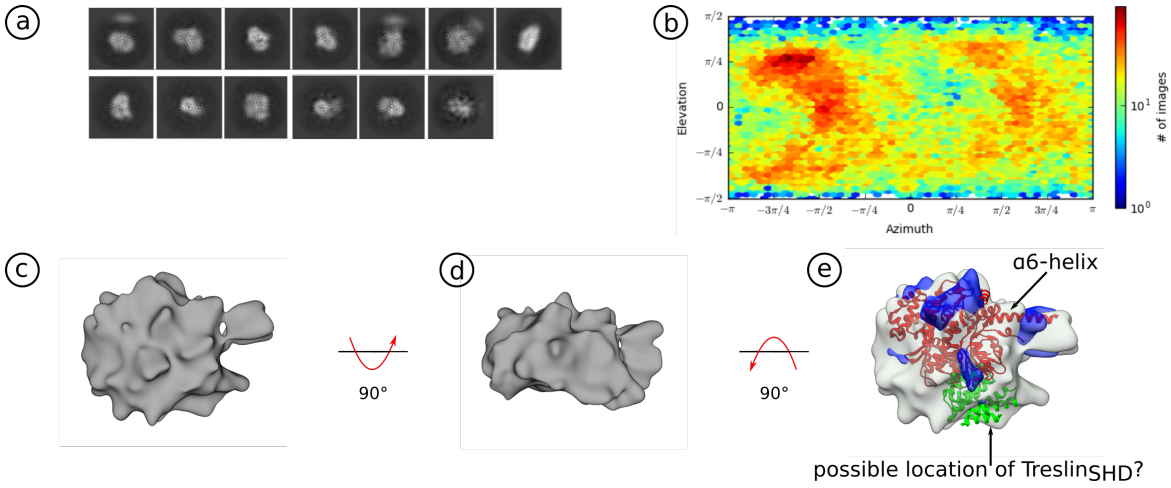



**Fig. 3.24 Cryo-EM analysis of the Cdc45–Treslin$_{SHD}$ complex with 0.005% (v/v) glutaraldehyde using cryoSPARC software package. (a)** Representative class averages, ranked by the abundance of particles, after 2D classification. **(b)** Angular distribution of the resulting 3D map. **(c–d)** 3D reconstructed map of Cdc45–Treslin$_{SHD}$ from cryoSPARC software package. **(e)** Fitting the crystal structure of human Cdc45 (highlighted in red, PDB: 5DGO) and Treslin$_{SHD}$ (highlighted in green, unpublished data) into the 3D density map (Simon et al., 2016). Blue blob represented Cdc45 structure by itself, obtained from figure 3.22.

RELION 3.0 software package to align the particles accurately for both 3D classification and auto-refinement steps. After fitting the crystal structure of Cdc45 to the map, it seemed that the signature $\alpha$-6 helix feature of the Cdc45 could not be fitted properly (figure 3.23g). The possibility of structural rearrangement of this $\alpha$-6 helix upon Treslin$_{SHD}$ binding could not be ruled out.

To verify and improve the map of Cdc45–Treslin$_{SHD}$ complex, the same dataset was re-processed using cryoSPARC 2.8.0 software package (Punjani et al., 2017). 339,783 particles from 1,592 motion-corrected micrographs were automatically picked and extracted using the cryoSPARC software package. After two consecutive rounds of 2D classification, 50,314 particles were selected based on their relative 'large' size which might correspond to the Cdc45–Treslin$_{SHD}$ complex (figure 3.24a). These particles were subsequently used for ab-initio model generation, asking for 3 classes. One well-resolved ab-initio model, which was larger than the other two (corresponding to 28,453 particles) were subsequently used for homogenous refinement step (figure 3.24b–d). The 3D reconstructed map was slight sharper and more interpretable than the one obtained from RELION 3.0's processing steps. The 3D map could accommodate both Cdc45 and Treslin$_{SHD}$ (figure 3.24e). However, at this low resolution, it is very difficult to unambiguously determine the precise location of both proteins.

Despite tremendous advancements in cryo-EM data acquisition and processing, dealing with small and dynamic protein complexes such as Cdc45–Treslin$_{SHD}$ or Cdc45–Treslin$_{SHD\text{-}ext}$ still posed a major challenge. The size of the protein complexes described in this chapter is between 87–115 kDa, which is smaller than larger macromolecular complexes solved previously using single-particle cryo-EM. There are structures of smaller proteins (<150 kDa) — such as haemoglobin, isocitrate dehydrogenase, lactate dehydrogenase, and streptavidin — that were elucidated by single-particle cryo-EM analysis. However, these analyses were aided tremendously by more stable samples, homo-oligomer composition of the protein complexes, and the availability of high-resolution X-ray crystallography structure for docking to the well-resolved 3D reconstructed map (Fan et al., 2019; Khoshouei et al., 2017; Merk et al., 2016).

### 3.2.13 Purification of full-length Treslin

Difficulties in obtaining the structure of Cdc45–Treslin$_{SHD}$ or Cdc45–Treslin$_{SHD\text{-}ext}$ by using single-particle cryo-EM analysis could be explained by the relatively small size of the complexes. Generating a larger and more stable protein assembly may result in well-behaved sample for cryo-EM analysis. One way to produce a bigger protein complex is by generating full-length Treslin (1910 amino acids, expected molecular weight = 210.8 kDa) in complex with Cdc45. Thus, full-length Treslin was cloned, over-expressed in either Sf9 insect or mammalian cells, and purified.

A full-length Treslin construct for over-expression in either Sf9 insects or HEK-293F mammalian cells was generated. The template containing the coding sequence was obtained from Professor William Dunphy of the California Institute of Technology (Caltech) in Pasadena, California, US (Kumagai et al., 2010). The protein-encoding gene was cloned into either a pACEMam1 (for mammalian cells expression) or pACEBac1 (for insect cells expression) vector, which also contains a human cytomegalovirus (CMV) promoter. A TEV-cleavable TwinStrep tag was added to the C-terminus. The rationale for this cloning strategy was to have an antibiotic-free constitutive promoter — compared to pCDNA5/TO that is inducible by tetracycline — and a specific affinity tag that could be utilised for downstream purification process.

The construct was transfected into 1.2 L of HEK-293F mammalian resuspension cells or 2 L of Sf9 insect cells. Unfortunately, the protein was expressed poorly and mostly insoluble regardless of the expression system (figure 3.25). Treslin may require the over-expression of its binding partner(s) to improve its expression level and solubility. Two binding partners of Treslin, Cdc45 and MCM, were co-expressed with Treslin. Unfortunately, co-expressing of

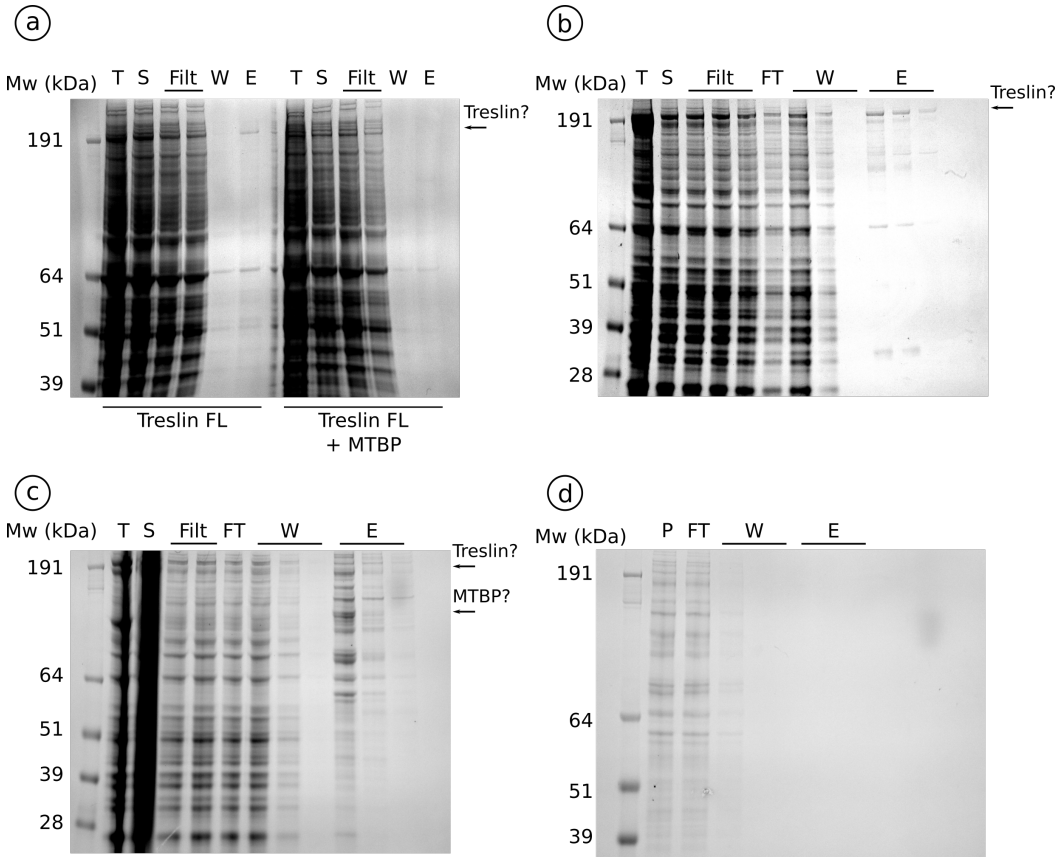



**Fig. 3.25 Expression of full-length Treslin and MTBP.** Coomassie-stained SDS-PAGE showing **(a)** initial StrepTactin capture of Treslin or Treslin–MTBP expressed in HEK-293F mammalian cells, **(b)** initial StrepTactin capture of Treslin produced in Sf9 insect cells, **(c)** initial Ni-NTA capture of Treslin–MTBP expressed in Sf9 insect cells, and **(d)** subsequent Strep capture of Treslin–MTBP from eluates in (d). Treslin was TwinStrep-tagged whereby MTBP was octahistidine-tagged. T = total, S = soluble, Filt = filtered lysate, FT = flow through, W = washes, E = eluates, P = pooled fraction. 4–12% gradient gels were used to run SDS-PAGE in these panels.

these proteins (individually or in combination) did not improve the yield and solubility of full-length Treslin.

MTBP (Mdm2-Binding Protein) was identified as an interacting partner of Treslin through work in *X. laevis* egg extracts (Boos et al., 2013). MTBP was subsequently suggested as the functional orthologue of budding yeast Sld7 in higher eukaryotes, as Sld7 was shown to interact with Sld3 throughout the cell cycle (Kumagai and Dunphy, 2017; Tanaka et al., 2011b). In a search of a way to boost the expression of Treslin, MTBP was subsequently co-expressed with Treslin in either Sf9 insect or HEK-293F mammalian cells.

Initially, full-length human MTBP was cloned into both pACEMam1 or pACEBac1 vectors with C-terminal hexahistidine tag. The coding sequence was obtained from the I.M.A.G.E clone 4476886 (SourceBioSciences). After co-expressing Treslin and MTBP in both expression systems, there was no improvement in terms of the yield and solubility of Treslin. Furthermore, the initial Ni-NTA affinity step captured most of the His-tagged MTBP but failed to elute Treslin. This observation demonstrated that this protein over-expression strategy failed to reconstitute the previously-reported Treslin–MTBP interaction (Boos et al., 2013).

## 3.3 Conclusions

In higher eukaryotes, Treslin acts as a loading factor for Cdc45 recruitment to the DNA replication origin. The interaction between Cdc45 and Treslin is a key step in regulating the assembly of CMG helicase (Boos et al., 2013; Kumagai et al., 2010). Despite its importance, there are very little biochemical and structural details regarding Cdc45–Treslin interaction and its regulation. Our current knowledge of Sld3/Treslin function is based largely on studies in model organisms such as *S. cerevisiae* or *X. laevis*. Despite the clear evolutionary relationship of the human system with the model organisms, human Treslin is a much larger protein that *S. cerevisiae* Sld3 (1910 vs 668 amino acids), and contains additional domains that are likely to confer extra layers of regulation and complexity to its function. Thus, one major aim of this project is to obtain mechanistic information about the formation, structure, and dynamic behaviour of the Cdc45–Treslin complex, which may provide a molecular basis for our understanding of Treslin's role in Cdc45 recruitment during the CMG helicase formation.

The complex of Cdc45–Treslin was reconstituted using purified recombinant proteins *in vitro*. It was verified by a number of pulldown experiments that a conserved homology domain of Treslin, termed Sld3/Treslin Homology Domain (SHD) or $\text{Treslin}_{\text{SHD}}$, is sufficient for binding with Cdc45. This result confirmed previous observations that showed the

SHD domain is a putative Cdc45-binding domain which is also well-conserved across the eukaryotes (Itou et al., 2014; Kumagai et al., 2010; Sanchez-Pulido et al., 2010). Through a combination of pulldown and semi-analytical size-exclusion chromatography experiments, it was found that Cdc45–Treslin complex formation is driven mainly by electrostatic interactions as higher salt condition abrogates complex formation. From the HDX-MS experiments, Cdc45's regions that interact with Treslin were also mapped. Treslin$_{\text{SHD-ext}}$ was shown to bind more extensively to Cdc45 than Treslin$_{\text{SHD}}$.

The C-terminal region next to the Treslin$_{\text{SHD}}$ is highly conserved in vertebrates. The sequence contains a number of positively-charged residues and potential S-CDK phosphorylation sites (T/S-P motif) (Errico et al., 2010). This particular domain, termed Treslin$_{\text{ext}}$, is predicted to be largely disordered. Two constructs representing Treslin$_{\text{SHD-ext}}$ and Treslin$_{\text{ext}}$ were generated, although their yield was much lower when compared to Treslin$_{\text{SHD}}$. Nevertheless, two constructs representing Treslin$_{\text{ext}}$ and Treslin$_{\text{SHD-ext}}$ could be purified to produce protein samples suitable for biochemical and biophysical characterisation.

Through pulldown experiments, it was shown that Treslin$_{\text{ext}}$ is able to interact with Cdc45 and this interaction is regulated by S-CDK phosphorylation. HDX-MS results showed that Treslin$_{\text{ext}}$ protected more regions of Cdc45 than Treslin$_{\text{SHD}}$, suggesting an extensive interaction of Cdc45 with Treslin$_{\text{ext}}$. These results strongly suggest the presence of additional Cdc45-binding site outside the Treslin$_{\text{SHD}}$, which is unique to the vertebrates. However, the interaction of the Cdc45–Treslin$_{\text{ext}}$ is weak as it was not maintained during semi-analytical size-exclusion chromatography. Furthermore, the Cdc45–Treslin$_{\text{ext}}$ interaction was abrogated by S-CDK phosphorylation which implies a role of S-CDK in negatively regulating the Cdc45–Treslin$_{\text{ext}}$ interaction.

SEC-MALS experiments showed that Cdc45 and Treslin$_{\text{SHD}}$ adopted a monomeric state in solution. In addition, Cdc45–Treslin$_{\text{SHD}}$ formed a near stoichiometric 1-to-1 complex in solution. The analysis involving the 'extension' region of Treslin was more complicated due to the disordered nature of the Treslin$_{\text{ext}}$ region. Treslin constructs containing the 'extension' domain (either Treslin$_{\text{SHD-ext}}$ or Treslin$_{\text{ext}}$) appeared to interact non-specifically with the column's resin and behaved unstably in low salt condition. This problem could be rectified when conducting the experiment in high salt buffer (300 mM NaCl) and/or S-CDK phosphorylation treatments of Treslin. The subsequent SEC-MALS experiments confirmed that each construct adopted a monomeric state in solution in the presence or absence of S-CDK phosphorylation.

High salt condition is not suitable for Cdc45–Treslin$_{\text{SHD-ext}}$ analysis as it inhibits complex formation. Therefore, the analysis has to be done in low salt condition (160 mM NaCl) which subsequently showed that Cdc45–Treslin$_{\text{SHD-ext}}$ formed a 1-to-1 complex in solution.

**Table 3.1** Binding affinity measurement for Cdc45–Treslin

| **Samples** | **Mean $K_d$ by BLI ($\mu$M)** | **Mean $K_d$ by SPR ($\mu$M)** | **Notes** |
|---|---|---|---|
| Cdc45–Treslin$_{SHD}$ | 4.8 | 8.3 | n/a |
| Cdc45–Treslin$_{SHD\text{-}ext}$ | 9.7 | 18.5 | no dissociation step in BLI |
| Cdc45–Treslin$_{ext}$ | 14.8 | cannot be determined | no dissociation step for both BLI & SPR |

Notes: $K_d$ = dissociation constant, BLI = Bio-Layer Interferometry, SPR = Surface Plasmon Resonance. All experiments were repeated at least twice

Treatment of Cdc45–Treslin$_{SHD\text{-}ext}$ with S-CDK resulted in a seemingly more stoichiometric complex that eluted as a sharp peak when compared to the non-phosphorylated Cdc45–Treslin$_{SHD\text{-}ext}$ complex. This could indicate that S-CDK promotes Cdc45–Treslin interaction, rather than inhibiting it, as observed in the pulldown experiments. However, it is likely that the SEC-MALS observation reflects an experimental artefact due to non-specific interactions of Treslin$_{SHD\text{-}ext}$ with the matrix of the gel filtration column, which then complicates the interpretation of the effect of S-CDK phosphorylation on the Cdc45–Treslin interaction.

Through Surface Plasmon Resonance (SPR) and Bio-Layer Interferometry (BLI), the binding affinity and the kinetics of Cdc45–Treslin complex were measured. The results are summarised in the table 3.1. It was noteworthy that the experiment involving Treslin$_{ext}$ produced a strange but reproducible phenomenon, in which the dissociation of the analyte (in this case, Cdc45) could not be observed even when changing buffer conditions or prolonging the dissociation time. This could be due to the limitation in the biosensor-based experiments or complex molecular reactions which could not be explained at our current understanding of Cdc45–Treslin complex. Due to the limited amount of Treslin$_{SHD\text{-}ext}$ or Treslin$_{ext}$, other measurements techniques that require a lot of purified proteins — such as Isothermal Titration Calorimetry (ITC) — could not be performed.

Structural analysis for Cdc45–Treslin proved to be a very challenging endeavour due to the small size of the complex and its dynamic nature, as observed by the BLI and SPR measurements. No crystal of the protein complex was observed during X-ray crystallography experiments despite extensive screening. Therefore, the protein complex was subjected to single-particle cryo-EM analysis. Due to the small size for each protein complex (87 kDa for Cdc45–Treslin$_{SHD}$ and 115 kDa for Cdc45–Treslin$_{SHD\text{-}ext}$), it was extremely challenging to obtain well-behaved samples after vitrification process and well-aligned particles during data processing.

Nevertheless, a low-resolution 3D reconstructed map of Cdc45 was obtained, in which the crystal structure of Cdc45 (PDB: 5DGO) could be fitted. No reliable evidence of Cdc45–Treslin structure was obtained by the 3D reconstruction, suggesting that the Cdc45–Treslin interaction was disrupted upon vitrification, likely due to air-water interface effects (Noble et al., 2018). A mild cross-linking treatment of the protein complex sample with glutaraldehyde suggested successful complex formation that survived the vitrification process, as judged by both 2D and 3D classification steps. However, the quality and quantity of the particles were not ideal for high-resolution 3D reconstruction. A low resolution map of Cdc45–Treslin$_{SHD}$ was obtained, in which an extra density was observed that represent Treslin$_{SHD}$. However, in such poorly resolved map, the location of both proteins could not be unambiguously assigned.

There is very little biochemical and structural information available on the full-length Treslin protein. Full-length Treslin was over-expressed in both insect and mammalian cells with an aim to characterise it biochemically and structurally either by itself and or in combination with Cdc45. Unfortunately, both expression systems failed to produce enough protein to perform any experiments. Over-expressing Treslin with other known binding partners, such as Cdc45 and MTBP, was also unsuccessful. There are potentially two ways for optimising Treslin's over-expression in the future. Firstly, a codon-optimised sequence of full-length Treslin for over-expression in mammalian cells could be used to boost the expression level. There is evidence that codon-optimisation of the sequence could increase the yield by factoring the availability of the tRNA, improving the rate of elongation, and increasing accuracy of translation (Kudla et al., 2009; Plotkin and Kudla, 2011). Secondly, Treslin may need to be expressed with other unknown protein partners in order to increase the yield. Unfortunately, within the time of this project, sufficient amount of pure full length Treslin could not be obtained.

In summary, the results presented in this chapter provide a biochemical, biophysical, and structural characterisation of purified Cdc45–Treslin complex in the context of DNA replication initiation. The findings show that the interaction of human Treslin with Cdc45 extends beyond the Treslin$_{SHD}$ to include a large, natively unstructured contiguous region C-terminal to the Treslin$_{SHD}$, referred to as 'extension' domain (Treslin$_{ext}$). This Treslin$_{ext}$ is capable of independent interaction with Cdc45 and contains several S-CDK sites, in addition to the known conserved S-CDK sites that bind to TopBP1 (T969 and S1001), that regulate the interaction with Cdc45. It is likely that these phosphorylation events mediate the chaperone function of Treslin in recruiting Cdc45 to the MCM during the CMG assembly process.

As a key factor in promoting assembly of the DNA-bound CMG helicase, Treslin might be expected to interact directly with the MCM proteins, as has been previously observed for

*S. cerevisiae* Sld3 (Deegan et al., 2016). A biochemical and structural investigation of the interaction of Treslin with MCM, TopBP1 (a known Treslin-binding substrate), and DNA are discussed in the next chapter.

# Chapter 4

# Treslin–Mcm2–7 interaction

## 4.1 Rationale

Mcm2–7 is an open ring protein complex that forms the core of the eukaryotic replisome. It is an apo-helicase that needs to be activated by two protein co-activators, monomeric Cdc45 and heterotetrameric GINS, to form an active replicative DNA holo-helicase called CMG complex (Gambus et al., 2006; Ilves et al., 2010; Moyer et al., 2006). CMG helicase assembly and activation result in the unwinding of duplex DNA, producing single-stranded DNA (ssDNA) templates for the replicative DNA polymerases. These highly regulated steps are the hallmark of DNA replication initiation in the eukaryotes.

Sld3/Treslin is crucial for orchestrating the stepwise recruitment of Cdc45 and GINS helicase co-activators to the DNA replication origin-bound Mcm2–7 (Boos et al., 2011; Kumagai et al., 2010; Sansam et al., 2010; Yeeles et al., 2015). In the *Saccharomyces cerevisiae* (budding yeast) system, the binding of Sld3 to Mcm2–7 is dependent on DDK (Dbf4-Dependent Cdc7 Kinase) phosphorylation of Mcm2, Mcm4, and Mcm6 subunits (Deegan et al., 2016; Heller et al., 2011; Herrera et al., 2015; Tanaka et al., 2011a). However, there is no report yet about direct Treslin–Mcm2–7 complex formation — and its regulation — in the human system.

This chapter describes the progress in characterising the interaction between Mcm2–7 and Treslin through a series of biochemical and biophysical experiments using purified recombinant human proteins. Soluble human Mcm2–7 was purified and its oligomeric state was analysed by Size-Exclusion Chromatography coupled with Multi-Angle Light Scattering (SEC-MALS). The purified Mcm2–7 sample was subjected to a single-particle cryo-electron microscopy (cryo-EM) analysis to obtain medium-resolution 3D reconstruction map.

Subsequently, Mcm2–7 interaction with Treslin was assessed in the absence or presence of DDK. In this setup, DDK-induced phosphorylation was not required for Treslin–Mcm2–7

complex formation, although a subset of purified Mcm2–7 was found to have been phosphorylated *in vivo*. However, the complex interaction was still observed even when Mcm2–7 was dephosphorylated by $\lambda$-phosphatase. Further experiments were able to map a region on Treslin that binds Mcm2–7 and uncovered a possible role of S-phase Cyclin-Dependent Kinase (S-CDK) in regulating Treslin–Mcm2–7 complex formation.

Apart from Mcm2–7- and Cdc45-binding properties (discussed in the previous chapter), $\text{Treslin}_{\text{ext}}$ (residues 813–1053) also possesses other biochemical characteristics involved in DNA replication initiation. $\text{Treslin}_{\text{ext}}$ was shown to interact with the N-terminal region of Topoisomerase II Binding Protein 1 (TopBP1) in a S-CDK-dependent manner. This indicated that this stage involving Treslin–TopBP1 during CMG assembly is conserved from fungi to human (Boos et al., 2011).

Furthermore, a novel DNA-binding domain in human Treslin was also discovered. The DNA-binding study made use of both Electrophoretic Mobility Shift Assays (EMSAs) and Fluorescence Anisotropy measurements. The results suggested that Treslin binds to both ssDNA and dsDNA equally well. Furthermore, the DNA-binding domain of Treslin has a preference for DNA sequences that form G-quadruplex (G4) secondary structures — a feature normally found around the DNA replication origin sequences (Prioleau, 2017). Taken together, the results provide a mechanistic insight into how Treslin is involved in DNA replication initiation through its ability to bind specific DNA structures.

The results presented in this chapter improve our understanding on how Treslin promotes the assembly of the active CMG helicase in human.

## 4.2 Results and Discussions

### 4.2.1 Purification of human Mcm2–7

The human Mcm2–7 complex was overexpressed in FreeStyle$^{\text{TM}}$ HEK-293F mammalian resuspension cells (ThermoFisher), via polyethylenimine (PEI) transient transfection. There are two major reasons for using HEK-293F cells. Firstly, codon-optimised DNA constructs that were suitable for overexpression of Mcm2–7 in mammalian cell culture had been developed in the laboratory before this project started. Secondly, the use of a human-derived cell line such as HEK-293F may promote appropriate *in vivo* post-translational modifications (such as phosphorylation) of the exogenous gene product.

The purification of human Mcm2–7 took advantage of two affinity tags that were engineered in the DNA coding sequences: octahistidine at the N-terminus of Mcm4 and TwinStrep at the N-terminus of either Mcm2 or Mcm5. These two tags helped to purify the

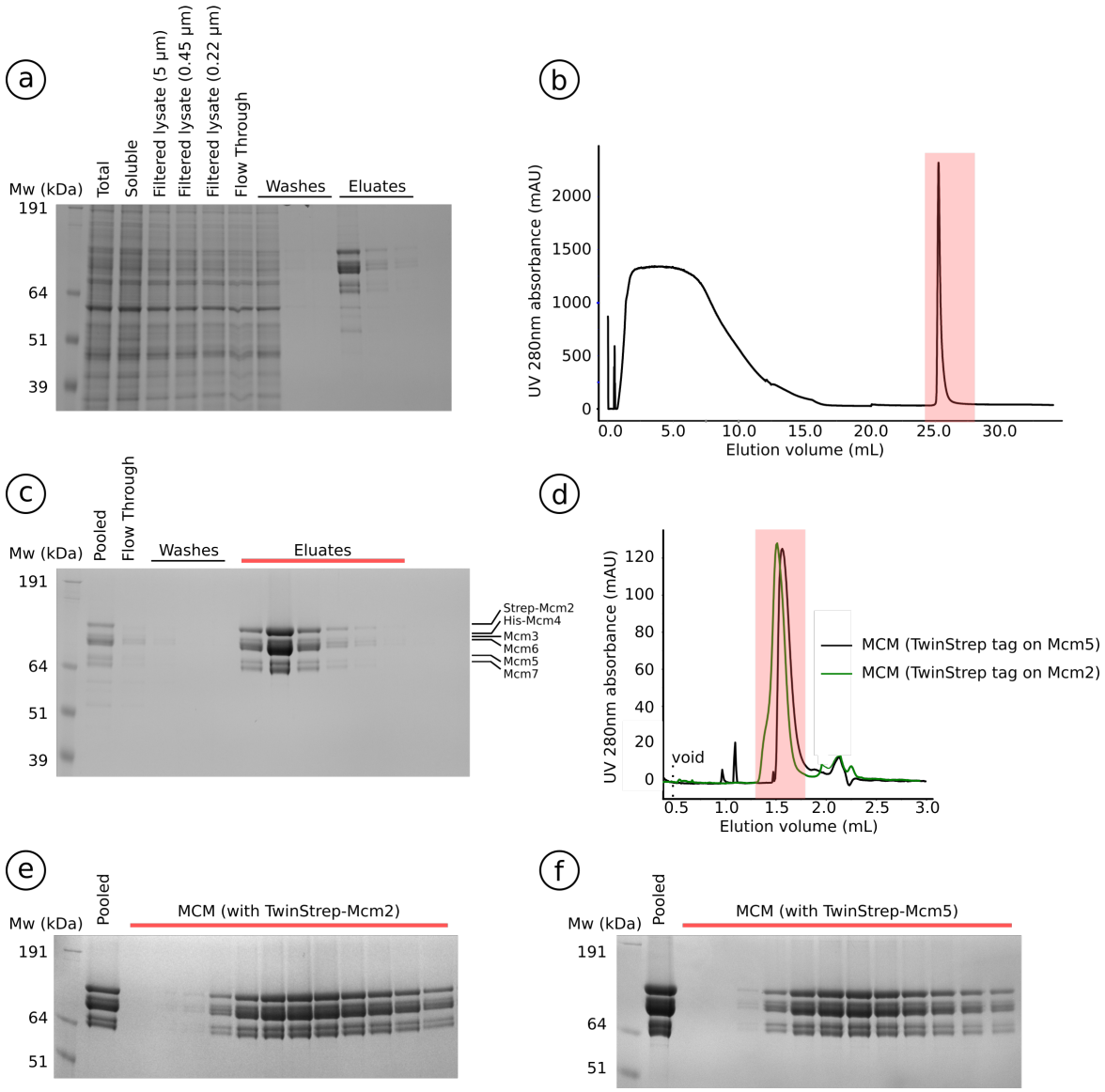



**Fig. 4.1 Purification of Mcm2–7.** **(a)** Coomassie-stained SDS-PAGE showing Ni-NTA affinity chromatography of Mcm2–7. **(b)** Chromatogram of subsequent StrepTactin affinity chromatography of Mcm2–7. **(c)** Coomassie-stained SDS-PAGE corresponding to (c). **(d)** Chromatogram of size-exclusion chromatography run using Superose6 Increase 3.2/300 run of Mcm2–7 with a TwinStrep tag either on Mcm2 (green) or Mcm5 (black). **(e)** Coomassie-stained SDS-PAGE of Mcm2–7 with a TwinStrep tag on Mcm2 based on run in (d). **(f)** Coomassie-stained SDS-PAGE of Mcm2–7 with a TwinStrep tag on Mcm5 based on run in (d).

hexameric Mcm2–7 from the potential contaminants of Mcm4-6-7 (Ma et al., 2010; Sato et al., 2000). The clarified lysate was also treated with benzonase nuclease to remove DNA and RNA contamination. In short, the purification steps of human Mcm2–7 hexamer are Ni-NTA followed by StrepTactin affinity chromatography. These steps produced pure and soluble Mcm2–7 complex, as judged by Coomassie-stained SDS-PAGE (figure 4.1a–c).

Purified and concentrated Mcm2–7 sample (at 1 mg/mL) for either construct were analysed by semi-analytical size-exclusion chromatography using a Superose6 Increase 3.2/300 column (GE Healthcare). Both constructs produced a main peak at 1.5 mL (figure 4.1d–f). Interestingly, an early 'shoulder' peak was also observed for the Mcm2–7 sample with a TwinStrep tag on Mcm2 (figure 4.1d, green line). It is possible that this early 'shoulder' peak represented a higher molecular weight species corresponding to Mcm2–7 double-hexamer (MCM-DH) species that could not be fully resolved on this column.

To confirm the oligomeric state of both Mcm2–7 constructs in solution, samples were analysed by SEC-MALS. Purified Mcm2–7 of either construct at 1 mg/mL were applied to a Superose6 Increase 3.2/300 size-exclusion chromatography column. Mcm2–7 with a TwinStrep tag on Mcm5 was eluted in a single peak of 603.0 kDa ($\pm$0.3 kDa), close to its predicted hexameric size of 558.6 kDa (figure 4.2a). This result proved that the Mcm2–7 construct with a TwinStrep tag on Mcm5 existed as an obligate single hexamer in solution. Meanwhile, the Mcm2–7 sample with a TwinStrep tag on Mcm2 eluted in a peak with a pronounced early 'shoulder', as expected from the previous semi-analytical size-exclusion chromatography experiment. The highest peak was analysed and showed a molecular weight of 669.4 kDa ($\pm$0.3 kDa), close to its predicted hexameric size of 558.6 kDa (figure 4.2b). When the early 'shoulder' was analysed separately, it revealed increasing molecular weight that plateaued at 1.24 MDa ($\pm$0.6 kDa) with polydispersity value of 1.006. It was possible that this shoulder represented Mcm2–7 that adopts a double-hexamer conformation, since 1.24 MDa is roughly twice the molecular weight of a single hexamer ($2 \times 669.4$ kDa). It was also likely that the higher molecular weight species were aggregated proteins, although this argument was undermined by the fact that the elution point for the shoulder was long after the Superose6 Increase 3.2/300 column's void volume (0.8 mL, roughly a third of the total column volume).

While designing the Mcm2–7 over-expression construct, high-resolution structure of MCM-DH were not available yet. After the structure of the *S. cerevisiae* MCM-DH was published, it was revealed that the N-terminal tail of Mcm5, but not Mcm2, is involved in the inter-hexameric interaction that stabilises the double hexamer (Li et al., 2015; Noguchi et al., 2017). This might be an explanation why the Mcm2–7 construct with a TwinStrep tag on the N-terminus of Mcm5 was unable to adopt double hexamer structure, as shown

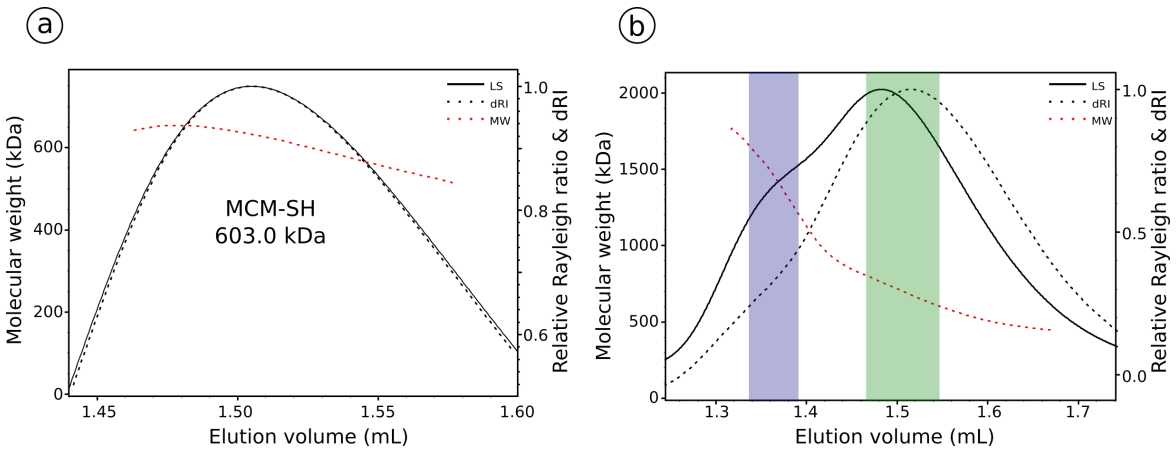



**Fig. 4.2 SEC-MALS of Mcm2–7. (a)** SEC-MALS of Mcm2–7 with a TwinStrep tag on Mcm5. Observed Mw = 603.0 kDa (±0.3 kDa), Polydispersity = 1.001 (±0.747%). **(b)** SEC-MALS of Mcm2–7 with a TwinStrep tag on Mcm2. 'Shoulder' peak (highlighted in blue): Observed Mw = 1.24 mDa (±0.6 kDa), Polydispersity = 1.006 (±0.466%). 'Highest' peak (highlighted in green): Observed Mw = 669.4 kDa (±0.3 kDa), Polydispersity = 1.001 (±0.421%). Expected Mw for MCM-SH = 558.6 kDa, expected Mw for MCM-DH = 1.1 MDa. See section 4.2.1 for more details

by the SEC-MALS result. Meanwhile, a very small proportion of Mcm2–7 proteins with a TwinStrep tag on N-terminus of Mcm2 was likely to adopt a double-hexamer conformation as observed during SEC-MALS. This was independent of DNA or other loading factors (figure 4.1). In summary, two different Mcm2–7 constructs were purified and characterised: one that is an obligate single hexamer (with a TwinStrep tag on Mcm5) and another that has the potential to form a double-hexamer (with a TwinStrep tag on Mcm2).

### 4.2.2 Cryo-EM analysis of Mcm2–7

#### Apo-Mcm2–7

Our structural understanding of eukaryotic Mcm2–7 came from *S. cerevisiae* or *D. melanogaster* proteins, as no structure of human Mcm2–7 is available (Li et al., 2015; Noguchi et al., 2017; Zhai et al., 2017a). To fill this gap, structural analysis of Mcm2–7 was performed using single-particle cryo-electron microscopy (cryo-EM). The purified Mcm2–7 sample, with a TwinStrep tag on Mcm2, was applied to a Superose6 Increase 3.2/300 size-exclusion chromatography column, as described earlier (figure 4.1e). The peak fraction, at 1 $\mu$M, was applied to a glow discharged R1.2/1.3 copper grid with a thin layer of carbon film (QuantiFoil). After blotting the excess liquid, the grid was rapidly vitrified in liquid ethane to embed the sample in vitreous ice. The grid was used for image acquisition using a Titan

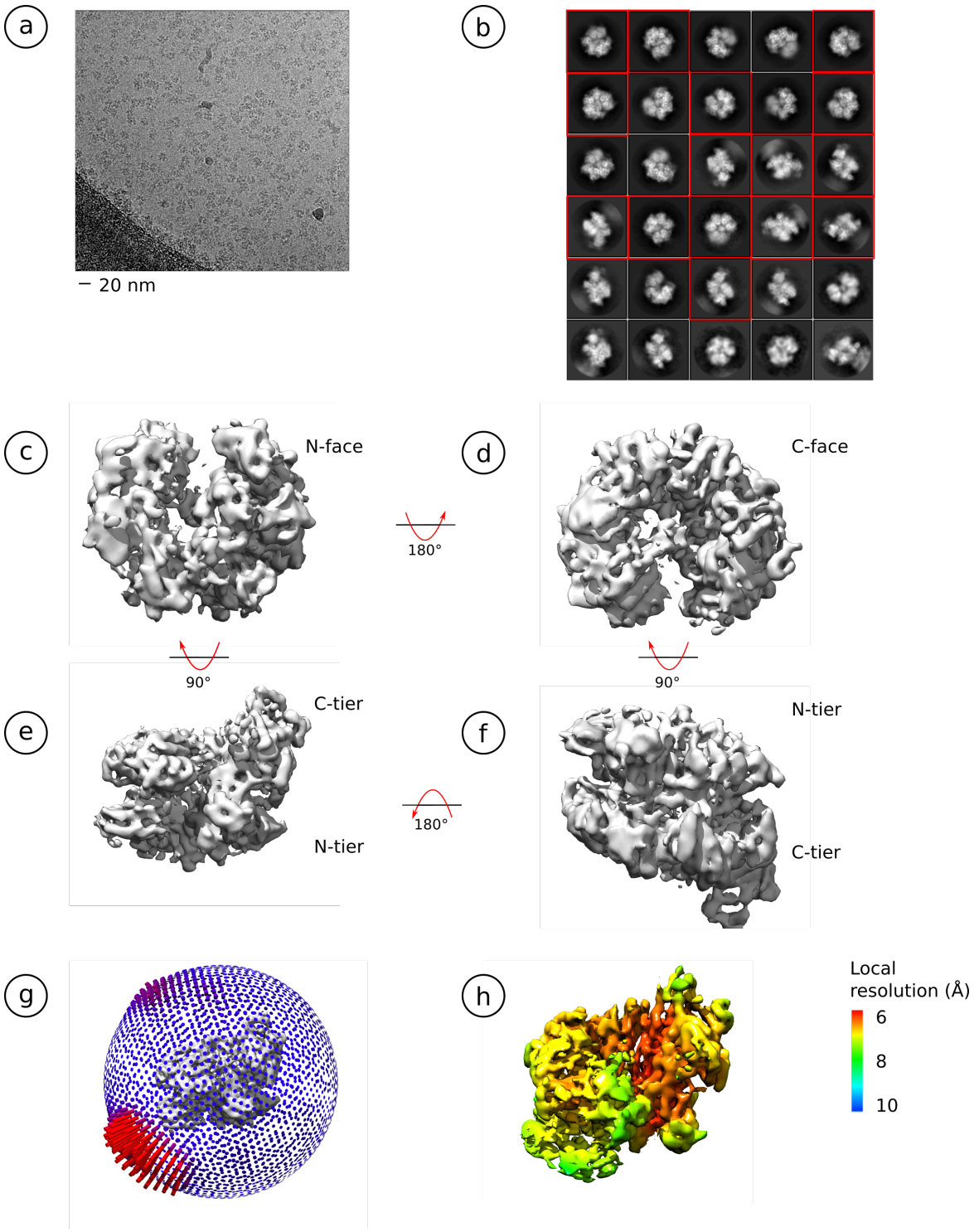



**Fig. 4.3 Cryo-EM analysis of Mcm2–7.** **(a)** Representative micrograph of the Mcm2–7 specimen. **(b)** 2D class averages of Mcm2–7. The classes selected for subsequent 3D reconstruction steps are highlighted in red boxes. **(c)–(f)** 3D reconstruction of Mcm2–7. **(g)** Angular distribution of the 3D map shown in c-f. **(h)** Local resolution of the 3D map of Mcm2–7, as determined in the RELION 2.1 software package (Scheres, 2016).

Krios microscope (FEI ThermoFisher) operated at 300 kV and equipped with a K2 Summit direct electron detector (Gatan Inc.) running in 'counting mode'.

644,719 particles from 3,500 motion-corrected micrographs were automatically picked and extracted using the RELION 2.1 software package (Scheres, 2016) (figure 4.3a). After three consecutive reference-free 2D classifications, distinct Mcm2–7 class averages could be observed (figure 4.3b). Mcm2–7 appeared to adopt an 'open' conformation, in which the 'gate' — likely to be in between the Mcm2 and Mcm5 subunits — was not locked down. Although each of the six MCM subunits could be clearly observed from the 2D class averages, at least two of the MCM subunits appeared to be poorly defined. This indicates that Mcm2–7 requires other substrates, such as other proteins or ATP, to stabilise the flexible subunits.

A subset of those class averages (13 classes, totalling 65,655 particles) was selected because they showed a relatively 'closed' conformation and less fuzzy 2D averages. The particles from these classes were subsequently used for 3D classification and auto-refinement in RELION 2.1 software, resulting in a 3D map with 6.8 Å resolution (figure 4.3h). The N-tier and C-tier of the Mcm2–7 single hexamer (MCM-SH) were clearly defined in the map. It adopted a shape akin to a 'lock-washer' (figure 4.3c–f). When compared to other Mcm2–7 structures in the context of either MCM double hexamer (MCM-DH) or CMG complex, MCM-SH is noticeably less planar (Li et al., 2015; Noguchi et al., 2017; Zhai et al., 2017a). A density was observed in the central channel of MCM-SH, suggesting the presence of a part of Mcm2–7 subunit that potentially prevents accidental insertion of dsDNA. It is also likely that the open gate is too narrow for dsDNA, as observed in *S. cerevisiae*'s MCM-SH (Zhai et al., 2017a).

The local resolution estimation of the structure highlighted the flexibility of Mcm2–7 sample (figure 4.3h). With respect to each other, the AAA+ ATPase in the C-tier was clearly more flexible compared to the N-tier. The density of the Mcm2–7 subunits that are away from the open gate was more clearly resolved (figure 4.3h, highlighted in red), as it is in a higher resolution compared to the two subunits that make up the gate. Unfortunately, in this resolution, it was not possible to measure the distance across the gate accurately. Nevertheless, it was likely that this MCM-SH was inactive and could not accommodate dsDNA binding possibly due to the small gap of the gate and presence of an MCM subunit's tail inside the central channel.

Docking of *S. cerevisiae*'s MCM-SH structure (PDB: 5H71) into the 3D map confirmed similarities between human and *S. cerevisiae* protein complexes (Zhai et al., 2017a). First, the density observed in the central channel of the MCM-SH corresponds to the Mcm5's C-terminal end (figure 4.4a, Mcm5 is coloured in red). The C-terminus of Mcm5 is likely to prevent accidental dsDNA insertion, as observed in *S. cerevisiae*. In addition, the open

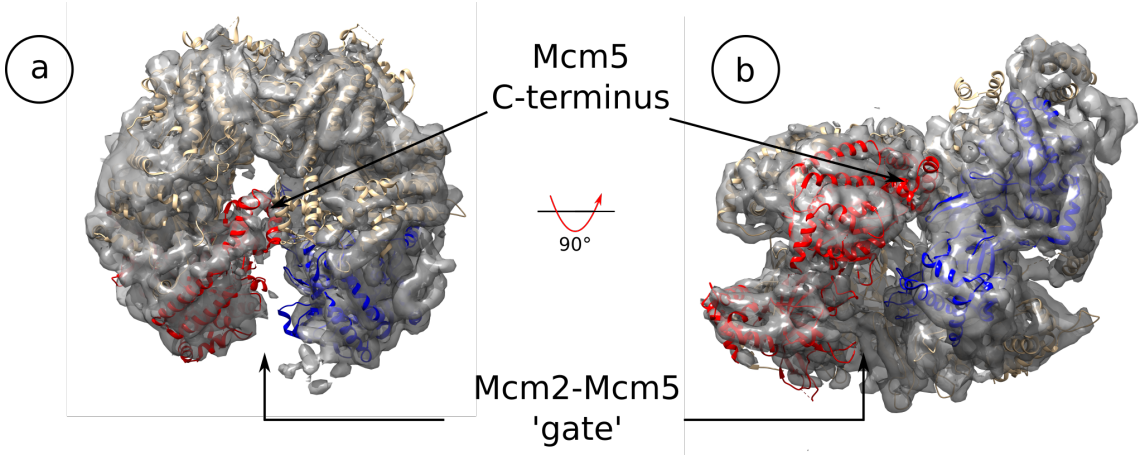



**Fig. 4.4 Docking of *S. cerevisiae*'s MCM-SH to the 3D map of human MCM-SH. (a)** The view from C-face **(b)** The side-view. Blue = Mcm2, Red = Mcm5. The gate and the C-terminus of Mcm5 are indicated with the arrows. PDB: 5H7I (Zhai et al., 2017a).

gate that exists between Mcm5 and Mcm2 (coloured in blue) is also too narrow for dsDNA insertion. Taken together, this structural feature of Mcm2–7 is also conserved across the eukaryotes.

There are several possible reasons why a high-resolution 3D map of Mcm2–7 could not be obtained. First, there was a problem of preferred orientation, as demonstrated by the angular distribution map (figure 4.3g). Second, Mcm2–7 construct used in this experiment contained a TwinStrep tag on Mcm2, with an assumption that this construct was the 'real' functional complex as it could adopt a double hexamer conformation (figure 4.2b). However, this inadvertently introduced heterogeneity in the sample (a lot of MCM-SH versus a few of MCM-DH) that complicated the analysis and could not be fully resolved during 2D or 3D classifications. Third, dense particle distribution in the embedded ice might have caused problems during particle alignment, in which the extracted particles may have been too close to the neighbouring particles or contaminants (figure 4.3a and b). Lastly, another layer of heterogeneity was introduced by different degree of Mcm2–7 gate opening. As this sample was prepared without any addition of Mcm2–7's substrates such as ATP or DNA that can close the gate, Mcm2–7 could have infinitesimal change with respect to the degree of opening of the Mcm2–7 gate. This was clearly demonstrated in the 2D classification results (figure 4.3b).

### Mcm2–7 + ATP

In an attempt to solve some of the heterogeneity problems, the purified Mcm2–7 sample was incubated with 2 mM ATP for 1 h at 4 °C prior to application on the grid. The grid was

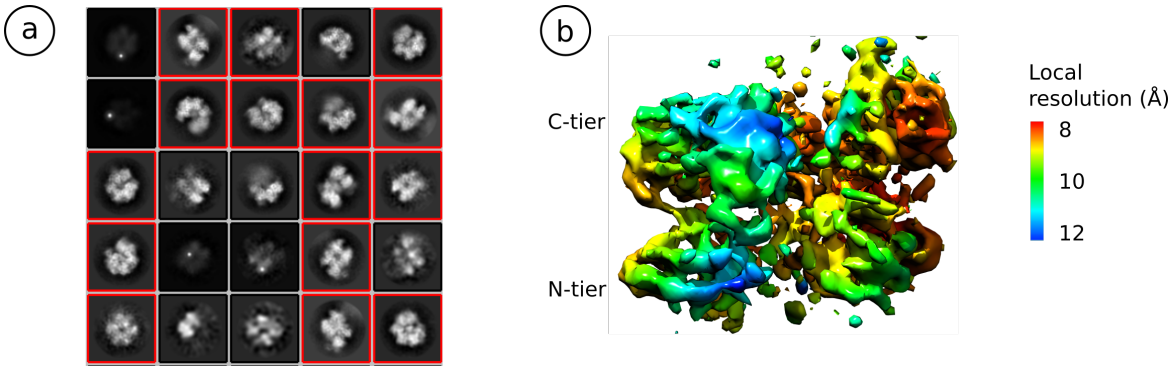



**Fig. 4.5 Cryo-EM analysis of Mcm2–7 + ATP. (a)** 2D class averages of Mcm2–7 + ATP. The classes selected for subsequent steps are highlighted in red boxes. The dark averages were indicative of thick vitreous ice on the grids. **(b)** 3D map of Mcm2–7 + ATP and local resolution estimated by locres in RELION 2.1 software.

used for data acquisition using a Titan Krios microscope (FEI ThermoFisher) operated at 300 kV and equipped with Falcon 3 direct electron detector (FEI ThermoFisher) running in 'linear' mode. Unfortunately, the vitreous ice that embedded the sample was too thick and this subsequently limited the downstream processing steps.

88,371 particles from the 1,320 motion-corrected micrographs were automatically picked. The result of 2D classification of these particles showed MCM-SH still adopting an 'open' ring conformation despite the addition of ATP. In addition, fuzzy density corresponding to at least two flexible MCM-SH subunits was still visible in the 2D class averages (figure 4.5a). These results suggested that addition of ATP to the Mcm2–7 was unable to lock the gate and/or stabilise the flexible Mcm2–7 subunits. Fifteen class averages (total particles = 25,300) were used to reconstruct 3D map. Unfortunately, the resolution was severely limited with thick ice present on the grid. Nevertheless, results from the 2D classification and the 3D reconstruction confirmed that the Mcm2–7 ring was still 'open' even in the presence of ATP (figure 4.5).

A few particles — likely to be the MCM-DH — were observed in a few of the micrographs. These species were manually picked but were eventually averaged out as the MCM-SH during the 2D classification (figure 4.3b, last row). There was limited 'side-view' of the MCM-DH, whereas the 'top-view' of the MCM-SH and MCM-DH is indistinguishable. Therefore, there was not enough particles to perform reliable 2D and 3D classification steps to perform 3D reconstruction for MCM-DH.

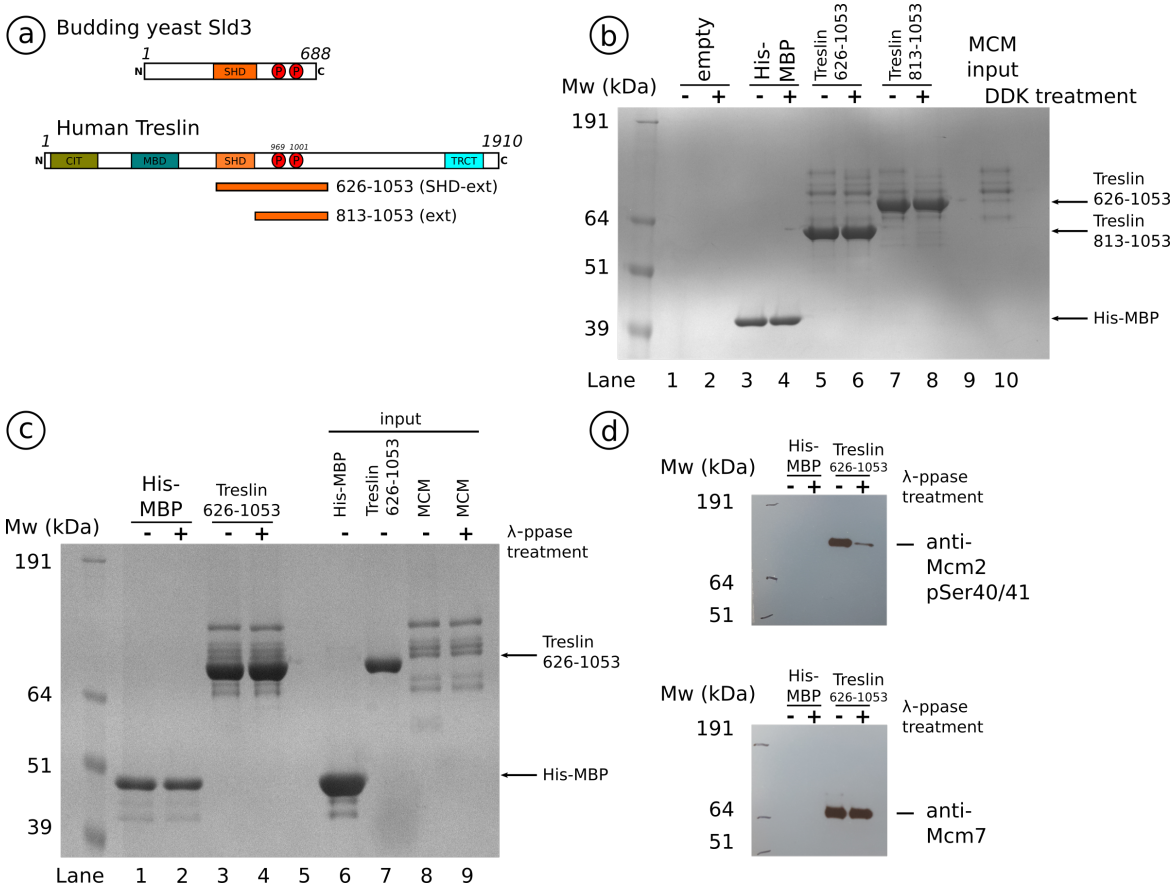



**Fig. 4.6 Pulldown experiment of MBP-tagged Treslin and Mcm2–7. (a)** Comparison between *S. cerevisiae* Sld3 and human Treslin. The MBP-tagged Treslin constructs used in this experiment are shown underneath. **(b)** Coomassie-stained SDS-PAGE showing the eluted fractions of amylose pulldown between MBP-tagged Treslin and Mcm2–7 in the absence or presence of DDK. **(c)** Coomassie-stained SDS-PAGE showing the eluted fractions of amylose pulldown between MBP-tagged Treslin and λ-phosphatase treated Mcm2–7. **(d)** Western blot result from protein samples lane 1–4 in panel (c) to show the effect of λ-phosphatase treatment using anti-Mcm2 pSer40/41 (top panel) and anti Mcm7 as a control (bottom panel).

### 4.2.3 Treslin–Mcm2–7 interaction

*S. cerevisiae* Sld3 was shown to form direct interaction with the Mcm2–7 ring, which offered a clue on Sld3/Treslin involvement during the CMG assembly step (Deegan et al., 2016). However, there is no evidence that the Mcm2–7-binding activity is also conserved in human Treslin.

An MBP pulldown experiment was performed to verify Treslin–Mcm2–7 interaction in the human system. Purified MBP-tagged constructs of $\text{Treslin}_{\text{SHD-ext}}$ and $\text{Treslin}_{\text{ext}}$ (residues 626–1053 and 813–1053, respectively) were used as 'bait' proteins to saturate the amylose agarose beads (figure 4.6a). The resin was washed thrice with buffer before adding the target protein (purified Mcm2–7 with a TwinStrep tag on Mcm2 expressed in HEK-293F mammalian cells). Before addition to the beads, Mcm2–7 was pre-phosphorylated with DDK. After washing the beads with buffer three times, the bound proteins were eluted with buffer containing 15 mM d-maltose and the sample was analysed by SDS-PAGE.

As shown in figure 4.6b, Mcm2–7 did not bind non-specifically to either the His-MBP control or the amylose resin. However, Mcm2–7 co-eluted from the beads in the case of MBP-tagged Treslin. Interestingly, Mcm2–7 co-eluted with MBP-tagged Treslin, regardless of the absence or presence of Mcm2–7's DDK treatment (as judged by the Coomassie-stained SDS-PAGE). Mcm2–7 also co-eluted with MBP-tagged $\text{Treslin}_{\text{ext}}$, which did not contain the SHD (figure 4.6b, lane 5 v 7 or 6 v 8). The result of this experiment shows that Mcm2–7 was able to form an interaction with Treslin via the the 'extension' domain ($\text{Treslin}_{\text{ext}}$) and independent from the SHD. This is in agreement with the previous finding in the *S. cerevisiae* system, in which Sld3 binds directly to the Mcm2–7, and suggests a conserved mechanism across eukaryotes (Deegan et al., 2016).

Intriguingly, in this experimental setup, DDK was not required for Treslin–Mcm2–7 interaction. It is very likely that Mcm2–7 was already phosphorylated *in vivo* during the expression in HEK-293F cells which primed Mcm2–7 for Treslin binding. To remove the *in vivo* phosphorylation, the Mcm2–7 sample was dephosphorylated — using the protocol suggested by the commercially available $\lambda$-phosphatase (SantaCruz Biotechnology, catalogue number sc-200312) — and subjected to another MBP pulldown experiment. The eluted proteins were analysed by SDS-PAGE. A similar amount of Mcm2–7 co-eluted from the beads with MBP-tagged $\text{Treslin}_{\text{SHD-ext}}$, regardless of $\lambda$-phosphatase treatment (figure 4.6c). $\lambda$-phosphatase was shown to be active, as judged by western blot analysis using a mouse antibody specific for phosphorylated human Mcm2 in serine residues 40 and 41 (figure 4.6d) (Montagnoli et al., 2006). This result suggested that the Treslin–Mcm2–7 interaction is independent from DDK in the human system, in contrast from what was observed in *S. cerevisiae* (Deegan et al., 2016). However, there is no detailed map of all phosphorylation

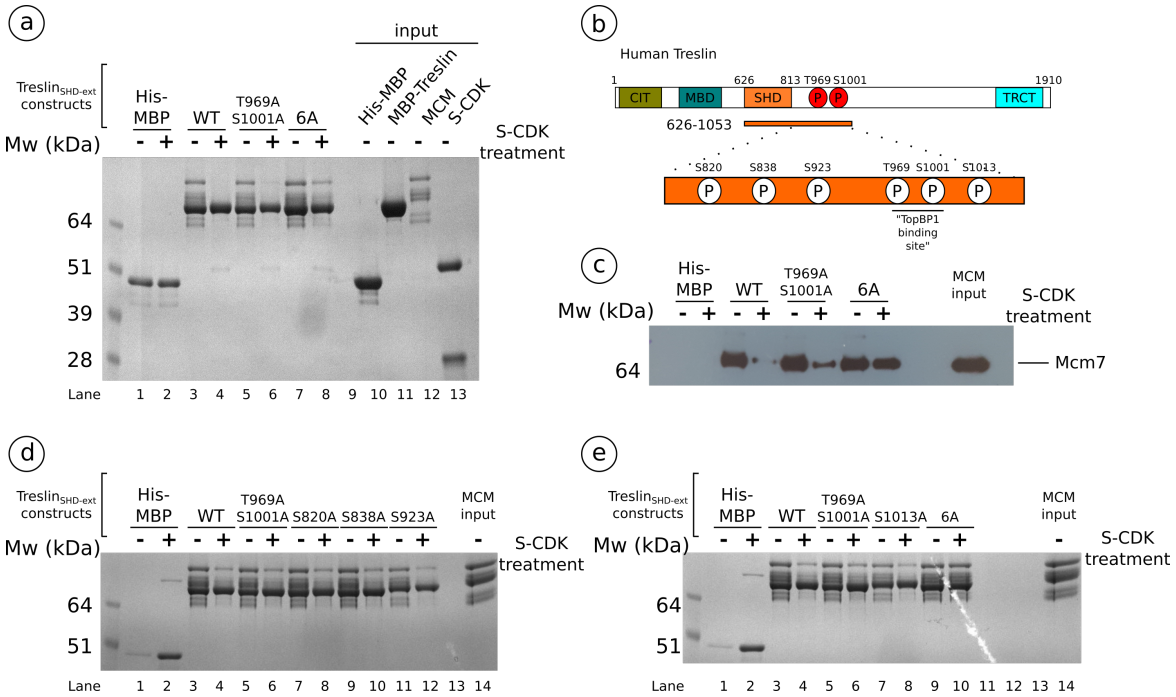



**Fig. 4.7 The effect of S-CDK on Mcm2–7 and Treslin interaction.** **(a)** Coomassie-stained SDS-PAGE showing the eluted fractions from MBP pulldown between MBP-tagged Treslin$_{\text{SHD-ext}}$ and Mcm2–7 in the absence or presence of S-CDK. **(b)** Schematic of six S-CDK sites that are mutated to alanine for the experiments described in this figure. **(c)** Western blot result from the eluted fractions in (a), detecting Mcm7. **(d)** Coomassie-stained SDS-PAGE showing the eluted fractions from MBP pulldown experiment between various MBP-tagged Treslin mutant constructs (S820A, S838A, S923A) and Mcm2–7. **(e)** As in (d), but using S1013A single mutant construct. WT = wild-type Treslin$_{\text{SHD-ext}}$, 6A = six alanine mutations on Treslin.

sites in human Mcm2–7 available to fully verify this observation. Nevertheless, for the subsequent experiments, the Treslin–Mcm2–7$_{\text{SHD-ext}}$ complex could simply be reconstituted without the need of DDK phosphorylation.

### 4.2.4 S-CDK regulates Treslin–Mcm2–7 interaction

As described in the previous chapter (chapter 3, section 3.2.5), the Treslin$_{\text{ext}}$ construct contains at least six potential S-CDK sites (serine/threonine-proline or S/T-P motifs). Two of these sites, threonine 969 and serine 1001 (T969 and S1001), were demonstrated to function as the TopBP1-binding site (Boos et al., 2011; Kumagai et al., 2011). The precise role of the other conserved S-CDK sites has not yet been determined. It is possible that S-CDK phosphorylation regulates Treslin's direct interaction with Mcm2–7, similar to what was observed with Cdc45 in the previous chapter.

To investigate the role of S-CDK phosphorylation on the Treslin–Mcm2–7 interaction, an MBP pulldown experiment was carried out. Purified MBP-tagged Treslin$_{\text{SHD-ext}}$ was used as 'bait' protein in the amylose agarose resin. In addition, two purified mutants of MBP-tagged Treslin$_{\text{SHD-ext}}$ were also used: one was a mutant carrying alanine mutations in residues S969 and T1001 (T969A + S1001A) and the other one was mutant carrying alanine mutations in all six potential S-CDK sites within Treslin$_{\text{SHD-ext}}$ (6A mutant which consists of S820A, S838A, S923A, T969A, S1001A, and S1013A). Each of the bound Treslin constructs was subjected to S-CDK phosphorylation treatment. After washing the beads three times with buffer to remove the S-CDK, purified Mcm2–7 was added. The bound proteins were eluted with buffer containing 15 mM d-maltose and analysed by SDS-PAGE and western blot (figure 4.7a and c). Mcm2–7 did not bind non-specifically to the free MBP tag (figure 4.7a, lane 1 and 2). As expected, Mcm2–7 co-eluted from the beads with non-phosphorylated MBP-Treslin$_{\text{SHD-ext}}$. Strikingly, less Mcm2–7 co-eluted with S-CDK-phosphorylated MBP-tagged Treslin$_{\text{SHD-ext}}$, indicating reduced Treslin$_{\text{SHD-ext}}$–Mcm2–7 complex formation (figure 4.7a, lane 3 v 4). This result demonstrates that the Mcm2–7 interaction with Treslin is abrogated in the event of Treslin phosphorylation by S-CDK, which provides a mechanistic clue of Treslin's disengagement from the replisome component Mcm2–7 prior to or during the DNA replication progression.

There was no observable difference when using the wild-type MBP-tagged Treslin$_{\text{SHD-ext}}$ and Treslin$_{\text{SHD-ext}}$ T969A + S1001A mutant — two known conserved Dpb11/TopBP1-binding sites in eukaryotes (figure 4.7a, lane 3–4 v 5–6). This suggested that these two sites alone are not responsible for S-CDK regulation and possibly only required for regulation of TopBP1 binding (Boos et al., 2011; Kumagai et al., 2011). To map which S-CDK site is responsible for regulating the Mcm2–7 and Treslin complex, for each single S-CDK site an alanine mutant construct was generated and used as a 'bait' protein for the MBP pulldown experiment (figure 4.7b). There was no single individual S-CDK site implicated on direct regulation of the complex stability (figure 4.7d and e). The disruption of the Treslin–Mcm2–7 interaction was less pronounced when using the MBP-tagged Treslin$_{\text{SHD-ext}}$ 6A mutant, which did not contain any potential S-CDK sites (figure 4.7e, lane 9 v 10). Taken together, these results demonstrated a role of S-CDK in regulating Treslin–Mcm2–7 complex stability, in which S-CDK phosphorylation abrogates direct interaction between Mcm2–7 and Treslin$_{\text{SHD-ext}}$. The two conserved sites for TopBP1 binding, T969 and S1001, are not responsible for this. Only when all the potential S-CDK sites were mutated to alanine, the complex of Treslin–Mcm2–7 was relieved from S-CDK-regulated disruption. This result suggests a degree of functional redundancy among the six S-CDK sites.

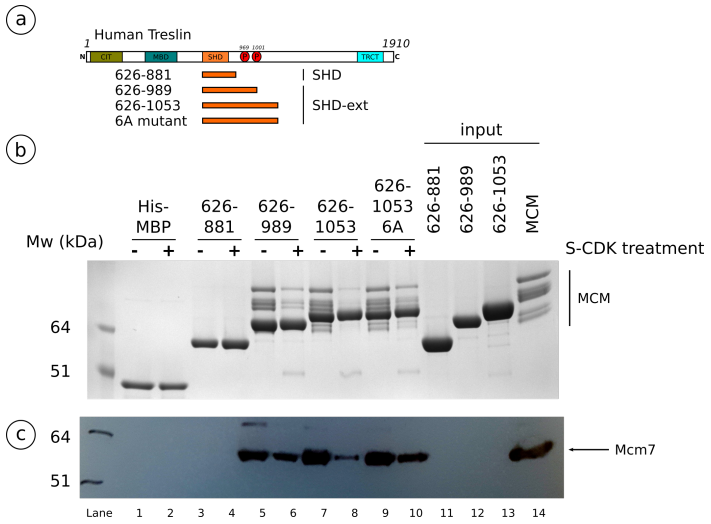



**Fig. 4.8 MBP pulldown of various Treslin constructs with Mcm2–7. (a)** Schematic of various MBP-tagged Treslin constructs used in this experiment. **(b)** Coomassie-stained SDS-PAGE showing the eluted fractions of MBP pulldown between various MBP-tagged Treslin and Mcm2–7 in the absence or presence of S-CDK. **(c)** Western blot result from (b) using anti-Mcm7 detection method.

To roughly map the region of Treslin that interacts with Mcm2–7, three truncated MBP-tagged Treslin constructs were generated. Two of them (corresponding to residues 626–989 and 626–1053) represented $\text{Treslin}_{\text{SHD-ext}}$, whereas residues 626–881 represented $\text{Treslin}_{\text{SHD}}$. MBP-tagged $\text{Treslin}_{\text{SHD-ext}}$ 6A mutant was also included in the experiment. All of these constructs were used as 'bait' proteins bound to the amylose agarose resin. After washing the beads with buffer for three times, the bound Treslin constructs were phosphorylated by S-CDK. Subsequently, purified Mcm2–7 was added into the reaction, and after washing the beads three times, bound proteins were eluted using buffer supplemented with 15 mM d-desthiobiotin and analysed by SDS-PAGE (figure 4.8b). Mcm2–7 co-eluted with all unphosphorylated MBP-tagged $\text{Treslin}_{\text{SHD-ext}}$ constructs, but not with MBP tag or MBP-tagged $\text{Treslin}_{\text{SHD}}$. Furthermore, less Mcm2–7 co-eluted with the S-CDK treated MBP-tagged $\text{Treslin}_{\text{SHD-ext}}$. This showed that region of $\text{Treslin}_{\text{ext}}$, between residues 882 and 1053, is sufficient to directly bind to the Mcm2–7.

The results from the MBP pulldown experiments were also observed when reversing the roles of 'bait' and 'target' proteins. Purified TwinStrep-tagged Mcm2–7 was used as the 'bait' protein in the StrepTactin agarose beads. An excess of purified and untagged $\text{Treslin}_{\text{SHD-ext}}$, pre-treated with S-CDK, was added to the reaction. After washing the beads three times, the bound proteins were eluted with buffer containing 15 mM d-desthiobiotin and analysed by SDS-PAGE (figure 4.9). Both Treslin constructs did not co-elute with the TwinStrep-tagged BirA control protein (BirA is an *E. coli* ligase that should not interact with Treslin),

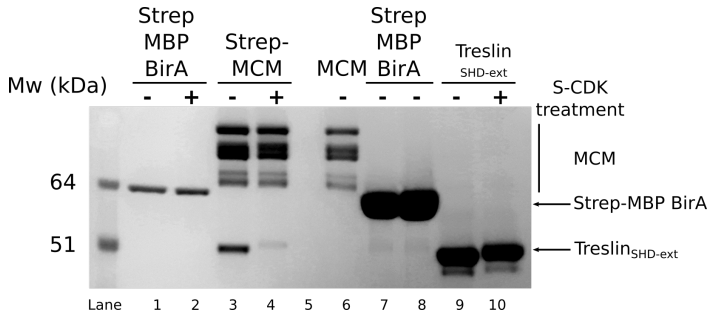



**Fig. 4.9 StrepTactin pulldown of Mcm2–7 and Treslin.** Coomassie-stained SDS-PAGE showing the eluted fractions of a StrepTactin pulldown between TwinStrep-tagged Mcm2–7 and untagged Treslin$_{\text{SHD-ext}}$ (residues 626–1053) in the absence or presence of S-CDK. TwinStrep-tagged MBP-BirA was used as a negative control.

indicating Treslin did not bind non-specifically to the TwinStrep tag. There was more of non-phosphorylated Treslin co-eluted with Mcm2–7 compared to the phosphorylated Treslin (figure 4.9, lane 3 v 4). This result confirmed the result of the MBP pulldown experiment, which showed that the Mcm2–7 interaction with Treslin is negatively regulated by S-CDK phosphorylation.

### 4.2.5 Treslin improves Cdc45 recruitment to the Mcm2–7

Studies in *S. cerevisiae* strongly suggested that Sld3 acts as a chaperone for Cdc45 recruitment to the Mcm2–7 complex (Deegan et al., 2016; Kamimura et al., 2001; Tanaka et al., 2011a). There is no biochemical evidence using purified human proteins that shows whether Treslin improves Cdc45 binding to Mcm2–7. If Sld3/Treslin is a functional chaperone and/or a loading factor for Cdc45, the presence of Treslin should improve the association of Cdc45 with Mcm2–7.

To test this hypothesis, a StrepTactin pulldown experiment was performed to reconstitute the Cdc45 loading reaction to the Mcm2–7. The first step was to use purified TwinStrep-tagged Mcm2–7 as the 'bait' protein for immobilisation in the StrepTactin agarose beads. Meanwhile, a complex of Treslin$_{\text{SHD-ext}}$ and Cdc45 in equimolar ratio was pre-formed by incubating them for 30 min on ice. Treslin$_{\text{SHD-ext}}$ was either treated or not treated with S-CDK before pre-forming of the complex. Afterwards, the beads were washed thrice and incubated together with pre-formed Cdc45–Treslin$_{\text{SHD-ext}}$ complex. The bound proteins were eluted with buffer containing 15 mM d-desthiobiotin and analysed with SDS-PAGE. Neither Treslin$_{\text{SHD-ext}}$, nor Cdc45, nor the pre-formed complex of Cdc45–Treslin$_{\text{SHD-ext}}$ bound non-specifically to the TwinStrep tag of the BirA control protein (figure 4.10, lane 1–6). There was a small amount of Cdc45 co-eluted with the Mcm2–7, regardless of the

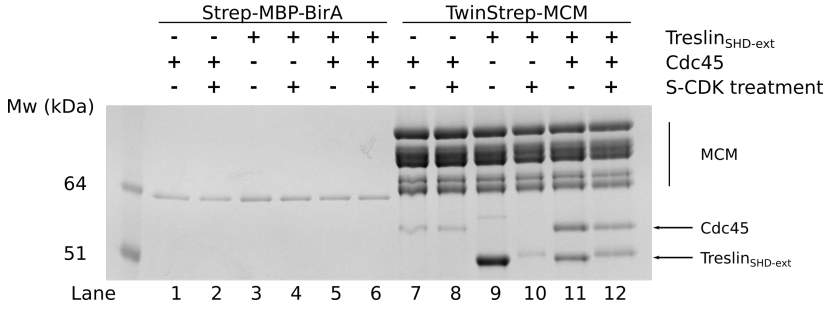



**Fig. 4.10 StrepTactin pulldown of Mcm2–7 with pre-formed Cdc45 and Treslin$_{SHD-ext}$ complex.** Coomassie-stained SDS-PAGE gel showing the eluted fractions of StrepTactin pulldown between TwinStrep-tagged Mcm2–7 with either Cdc45, Treslin$_{SHD-ext}$ (residues 626–1053), and pre-formed Cdc45–Treslin$_{SHD-ext}$ in the absence or presence of S-CDK.

S-CDK presence, which was indicated by weak band in the Coomassie-stained SDS-PAGE (figure 4.10, lane 7 and 8). There was more protein of unphosphorylated Treslin$_{SHD-ext}$ that co-eluted with Mcm2–7, when compared to the phosphorylated Treslin$_{SHD-ext}$ (figure 4.10, lane 9 v 10). These observations were expected as each of the individual proteins was a known interacting partner of Mcm2–7.

Interestingly, when Cdc45–Treslin$_{SHD-ext}$ complex was pre-formed in the absence of S-CDK phosphorylation, more Cdc45 co-eluted with the Mcm2–7 (figure 4.10, lane 7 v 11). This result showed that Treslin addition increased the amount of Cdc45 bound to the Mcm2–7. This data shows that Treslin is a loading factor for Cdc45, in which Treslin improves Cdc45 binding to the Mcm2–7. When the complex was pre-formed in the presence of S-CDK phosphorylation, less Cdc45 or Treslin$_{SHD-ext}$ co-eluted with Mcm2–7 (figure 4.10, compare lane 11 v 12). This further highlighted the mechanism of negative regulation by S-CDK phosphorylation for Treslin-dependent Cdc45 recruitment to the Mcm2–7.

### 4.2.6 Cryo-EM analysis of Treslin–Mcm2–7 complex

To obtain a Treslin–Mcm2–7 complex, a co-expression experiment was initially performed. Constructs of octahistidine-tagged Mcm2–7 and TwinStrep-tagged Treslin were co-expressed in HEK-293 mammalian cells. For this experiment, Mcm2–7 did not contain TwinStrep tag, in contrast with what was described earlier in this chapter (section 4.2.1). The rationale was to capture the complex via the octahistidine tag on Mcm4 using Ni-NTA affinity, followed by TwinStrep capture on Treslin. While Mcm2–7 expressed well, unfortunately Treslin did not express well as evidenced by a silver-stained SDS-PAGE (figure 4.11, right side). The same result was observed when Cdc45 was co-expressed with the other two constructs (figure 4.11, left side). Therefore, proto-CMG complex of either Treslin–Mcm2–7 or Cdc45–Treslin–

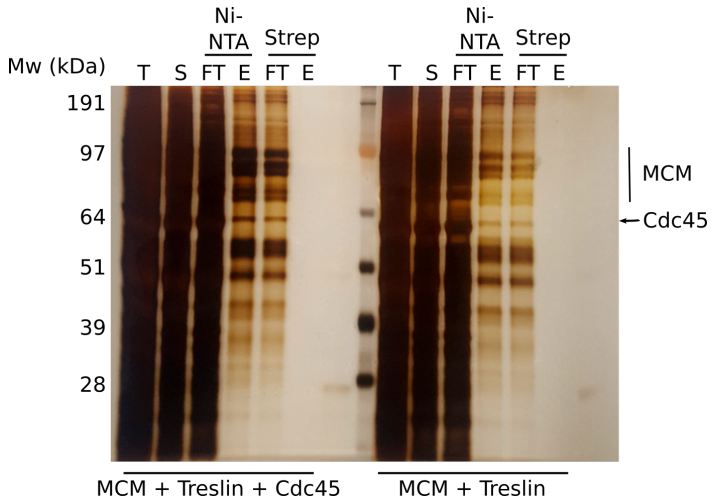



**Fig. 4.11 Co-expression trial of Mcm2–7-Treslin with or without Cdc45.** Silver-stained SDS-PAGE showing the purification attempts from co-expression experiments. Left side of the gel shows the co-expression of Cdc45-Mcm2–7-Treslin, whereas the right shows the co-expression of Treslin–Mcm2–7. T = total lysate, S = soluble fractions, FT = flow through, E = eluate, Ni-NTA = Nickel affinity step, Strep = StrepTactin affinity step.

Mcm2–7 could not be obtained via the co-expression method in HEK-293F mammalian cells.

An alternative way to obtain the Treslin–Mcm2–7 complex was to use the eluate from a pulldown experiment, as described earlier, using $\text{Treslin}_{\text{SHD-ext}}$ construct (figure 4.10, lane 9). The sample was subsequently embedded in vitreous ice by applying it to an UltrAuFoil R0.6/1.0 holey grid (UltrAuFoil) and plunge freezing the grid into liquid ethane. The grid was subsequently used for data acquisition in a Titan Krios microscope (FEI ThermoFisher) operated at 300 kV and equipped with a K2 Summit direct electron detector (Gatan Inc.) operated in 'counting'.

533,326 particles from 3,171 motion-corrected micrographs were automatically picked (figure 4.12a). After two consecutive reference-free 2D classifications, 81,152 particles from 20 satisfactory class averages were retained (figure 4.12b). After the 2D classification stage, most of the particles were classified to be Mcm2–7 by itself, without clear evidence of Treslin-bound Mcm2–7. However, there were some 2D averages not observed in the apo-Mcm2–7 analysis and they were highlighted with red arrow heads (compare figure 4.12b and 4.3b). These classes might contain additional density which could be accounted for $\text{Treslin}_{\text{SHD-ext}}$. After a round of 3D classification, a subset of these particles (44,300 particles, or 37.3% total of kept particles) were used to perform 3D reconstruction in RELION 3.0 software. The 3D map was subsequently refined to produce an 8 Å map (global resolution, based on RELION 3.0's locres estimation) (figure 4.12c–f and h). Unfortunately, the map

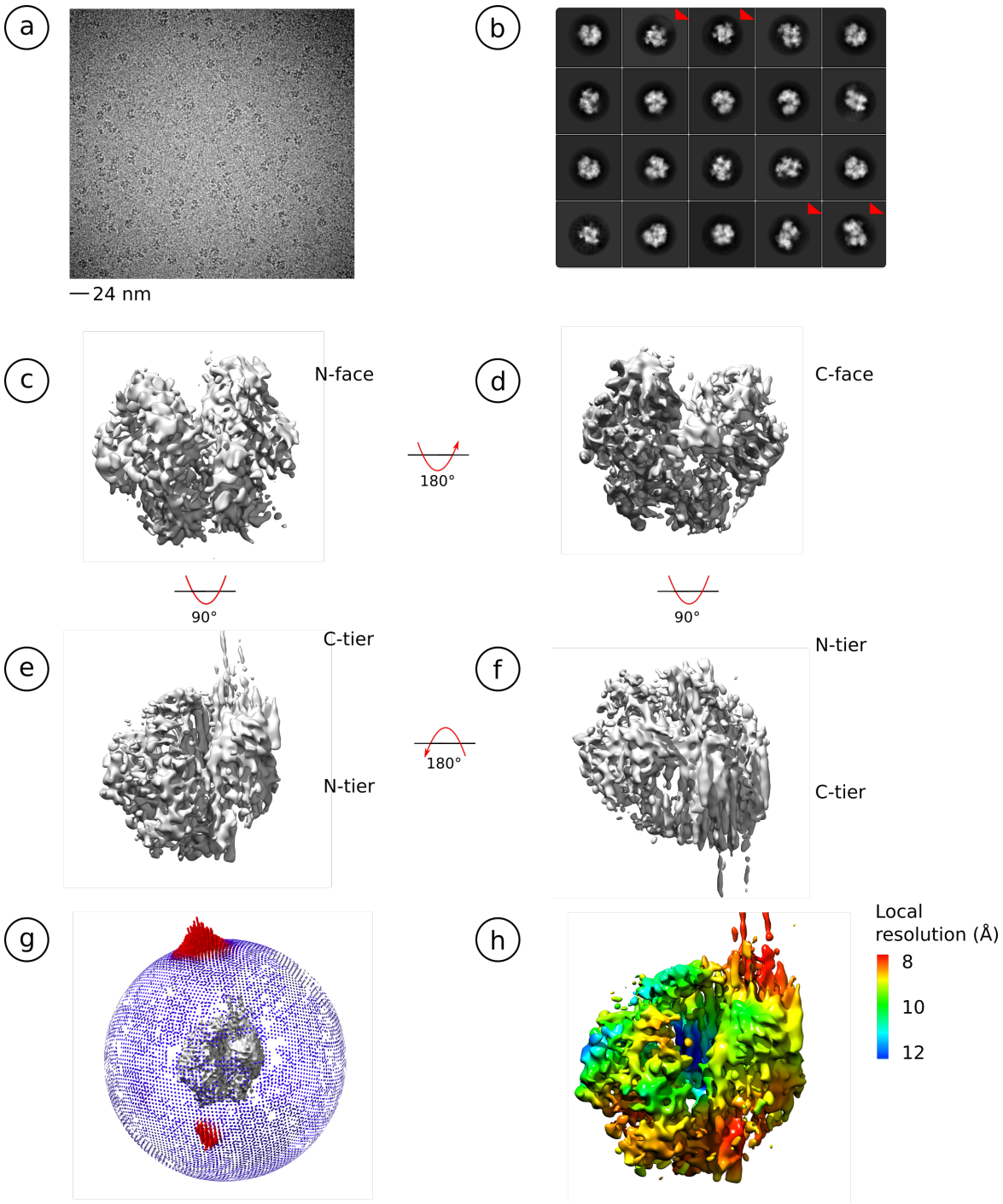



**Fig. 4.12 Cryo-EM analysis of Treslin$_{\text{SHD-ext}}$–Mcm2–7 complex.** **(a)** Representative micrograph of Treslin–Mcm2–7 specimen. **(b)** 2D class averages of Treslin–Mcm2–7. Red arrow-head represented 2D class averages that were not seen in apo-Mcm2–7 analysis. **(c)–(f)** 3D reconstruction of Treslin–Mcm2–7. **(g)** Angular distribution of the 3D map shown in c–f. **(h)** Local resolution of the 3D map of Treslin–Mcm2–7, as determined by locres in RELION 3.0 software (Scheres, 2016).

was very poorly resolved. This issue was likely to be caused by a severe problem of preferred orientation (figure 4.12g). It was likely that Treslin$_{\text{SHD-ext}}$–Mcm2–7 complex was disrupted upon freezing due to air-water interface effects (Noble et al., 2018). The sample was also inherently heterogenous, as it contained a mixture of Mcm2–7 (majority), Treslin$_{\text{SHD-ext}}$, and Treslin$_{\text{SHD-ext}}$–Mcm2–7 complex. Therefore, this 3D reconstruction failed to produce clear evidence of Treslin-bound Mcm2–7 after vitrification process.

### 4.2.7 Treslin interaction with TopBP1

Treslin residues of threonine 969 (T969) and serine 1001 (S1001) are conserved across the eukaryotes. Upon S-CDK phosphorylation, these two sites serve as the binding site for the N-terminal BRCT repeats of Dpb11/TopBP1 which has been shown both in budding yeast and human systems (Boos et al., 2011; Kumagai et al., 2011; Zegerman and Diffley, 2007). This conserved interaction between TopBP1–Treslin in humans (Dpb11–Sld3 in fungi) is crucial for downstream recruitment process of GINS and Pol$\varepsilon$ prior to DNA unwinding and synthesis (Muramatsu et al., 2010).

Using recombinant human proteins, Treslin–TopBP1 complex formation was reconstituted using an MBP pulldown experiment. Three different MBP-tagged Treslin constructs — Treslin$_{\text{SHD}}$, Treslin$_{\text{SHD-ext}}$, and Treslin$_{\text{ext}}$ — were used as the 'bait' protein to saturate the amylose agarose resin (figure 4.13a, top three constructs). The beads were washed three times with buffer to remove excess 'bait' protein. Afterwards, the bound MBP-tagged Treslin constructs were phosphorylated by S-CDK. After washing the beads thrice with buffer, a two-fold molar excess of purified N-terminal domain of TopBP1 (TopBP1 NTD) was added. The bound proteins were eluted by buffer containing 15 mM d-maltose and analysed by SDS-PAGE (figure 4.13b). TopBP1 NTD bound neither to the amylose beads nor the MBP tag. TopBP1 NTD did not co-elute with MBP-tagged Treslin$_{\text{SHD}}$ (figure 4.13 lane 5 and 6), which was expected since this particular Treslin construct did not contain T969 and S1001 residues. TopBP1 NTD did not co-elute with either non-phosphorylated MBP-tagged Treslin$_{\text{SHD-ext}}$ or Treslin$_{\text{ext}}$. However, TopBP1 NTD co-eluted with phosphorylated MBP-tagged Treslin$_{\text{SHD-ext}}$ and Treslin$_{\text{ext}}$ (figure 4.13, lane 7 v 8 and 9 v 10). This result showed that the NTD of human TopBP1 is sufficient to interact with phosphorylated Treslin$_{\text{SHD-ext}}$ or Treslin$_{\text{ext}}$. These two constructs contain the conserved residues of T969 and S1001. Taken together, these results confirmed the conserved mechanism of TopBP1–Treslin binding across eukaryotes, in which its formation is dependent on S-CDK phosphorylation.

A phosphomimetic mutant of Treslin$_{\text{SHD-ext}}$ was generated to test whether the S-CDK requirement can be bypassed for Treslin–TopBP1 complex formation. The initial approach was to over-express a mutated construct that contains amber codon mutations in an *E. coli*

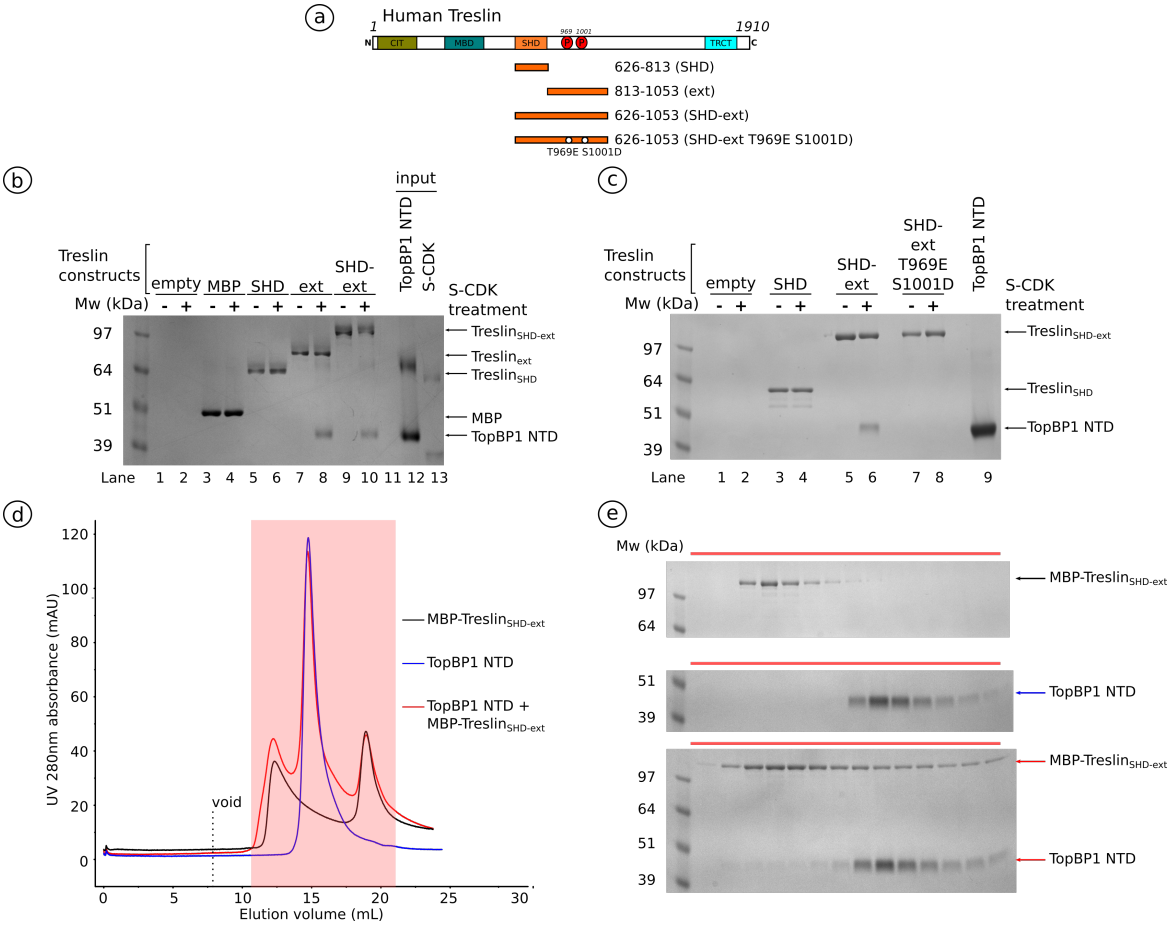



**Fig. 4.13 Treslin interaction with N-terminal region of TopBP1 (TopBP1 NTD). (a)** Schematic of MBP-tagged Treslin constructs used in these experiments. **(b)** Coomassie-stained SDS-PAGE showing the eluted fractions from MBP pulldown between various MBP-tagged Treslin and TopBP1 NTD in the absence or presence of S-CDK. **(c)** Coomassie-stained SDS-PAGE showing the eluted fractions from MBP pulldown experiment between MBP-tagged Treslin$_{\text{SHD-ext}}$ phosphomimetic mutant and TopBP1 NTD in the absence or presence of S-CDK. **(d)** Chromatogram of semi-analytical size-exclusion chromatography experiment using Superdex200 10/300 column of MBP-tagged Treslin$_{\text{SHD-ext}}$ and TopBP1 NTD in the presence of S-CDK. **(e)** Coomassie-stained SDS-PAGE showing the eluted fractions from the peaks highlighted in (d).

strain with modified tRNA/tRNA synthase system to incorporate phosphoserine while reading the 'amber' (TAG) codon (Rogerson et al., 2015). Unfortunately, this method failed to yield enough soluble protein for any biochemical experiment. An alternative approach was to mutate T969 to glutamic acid (T969E, to mimic phosphothreonine) and S1001 to aspartic acid (S1001D, to imitate phosphoserine) to produce MBP-tagged $\text{Treslin}_{\text{SHD-ext}}$ T969E + S1001D mutant (figure 4.13a, bottom construct). The resulting MBP-tagged phosphomimetic $\text{Treslin}_{\text{SHD-ext}}$ T969E + S1001D was used as bait protein and bound to the amylose agarose resin for another MBP pull-down experiment. Interestingly, TopBP1 NTD did not co-elute with MBP-tagged Treslin T969E + S1001D mutant, with or without S-CDK phosphorylation treatment (figure 4.13c, lane 7 and 8). This result showed that phosphomimetic Treslin construct failed to bind to TopBP1 NTD, which suggests the specificity of phosphoserine and phosphothreonine in recognising the BRCT repeats in TopBP1 NTD.

To reconstitute and characterise the complex of $\text{Treslin}_{\text{SHD-ext}}$ and TopBP1 NTD, a semi-analytical size-exclusion chromatography experiment was performed. Purified MBP-tagged $\text{Treslin}_{\text{SHD-ext}}$ at 50 $\mu$M was phosphorylated by S-CDK before incubating with purified TopBP1 NTD at 100 $\mu$M. The mixture was applied to a Superdex200 10/300 semi-analytical size-exclusion chromatography column (figure 4.13d). Each of the sample corresponding to the peaks were analysed by SDS-PAGE (figure 4.13e). There was no higher molecular species corresponding to the size of MBP-tagged $\text{Treslin}_{\text{SHD-ext}}$ and TopBP1 NTD (figure 4.13e, bottom panel). This suggested that the complex mixture fell apart and could not be sustained during the size-exclusion chromatography run.

### 4.2.8 Treslin contains a DNA-binding domain

DNA contamination — a common problem for DNA-binding proteins — was observed during purification of $\text{Treslin}_{\text{SHD-ext}}$, but not of $\text{Treslin}_{\text{SHD}}$. The contaminant could only be removed by a heparin affinity chromatography step (chapter 3). Since there was no published report of DNA-binding ability of Treslin, it was interesting to investigate whether Treslin contains a novel DNA-binding domain.

Initially, an Electrophoretic Mobility Shift Assay (EMSA) experiment was performed using purified $\text{Treslin}_{\text{SHD-ext}}$ or $\text{Treslin}_{\text{ext}}$ with dsDNA. The DNA oligonucleotide is labelled by 6-FAM at the $5'$-end and consisted of random 20 nucleotide sequences (figure 4.14a). A 20-mer DNA was used because multiple copies of Treslin could bind to a longer piece (>30 bp) of dsDNA (data not shown). In a non-denaturing agarose gel, fluorescently-labelled free DNA should migrate faster compared to protein-bound DNA. The change in migration distance of fluorescently-labelled DNA can be interpreted as a protein–DNA interaction. The binding affinity ($K_d$) of the protein for the DNA can be roughly determined by the

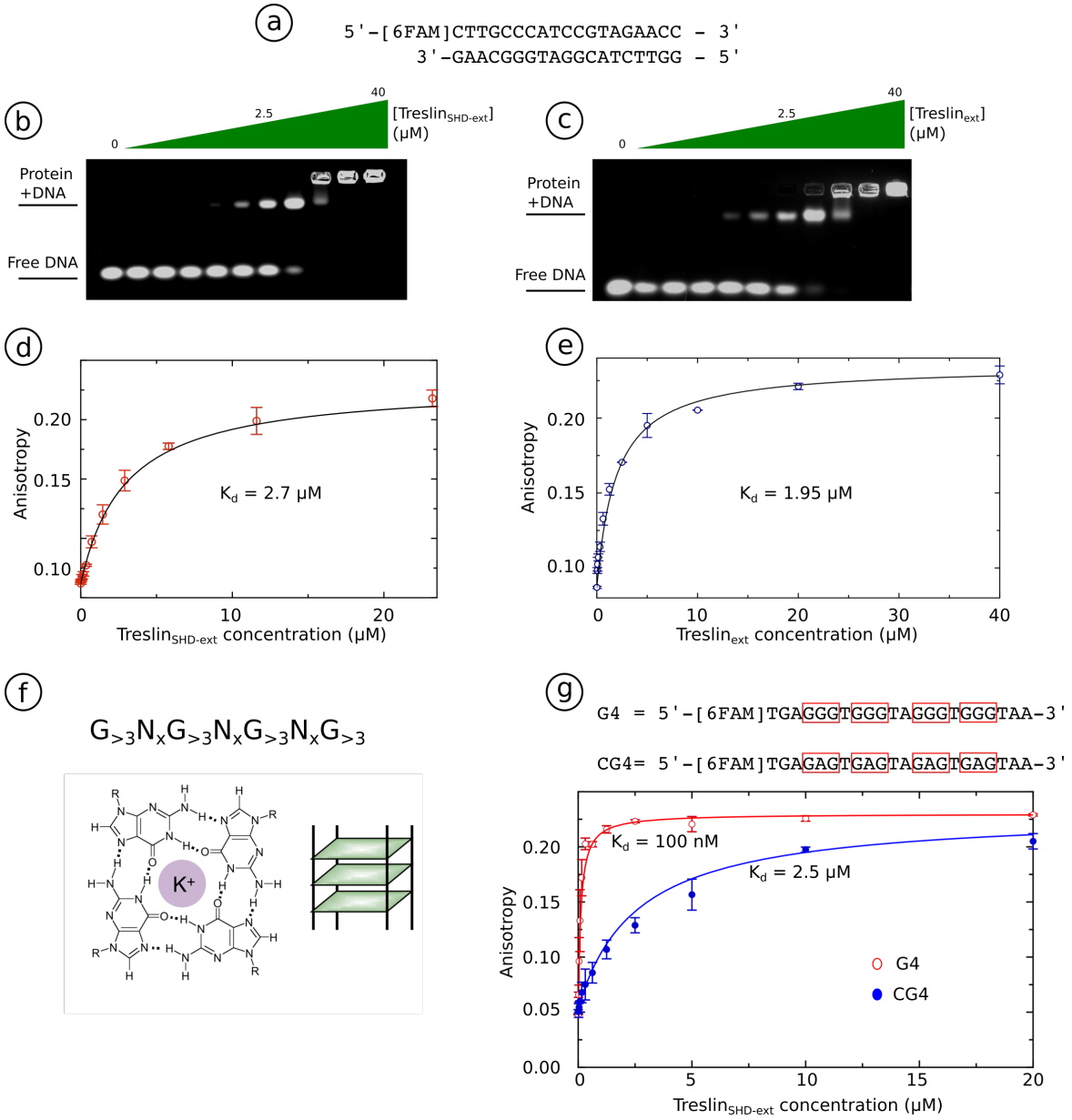



**Fig. 4.14 Treslin interaction with DNA. (a)** A double-stranded DNA (dsDNA) sequence used in the experiments shown in (b)-(e). **(b)** Electrophoretic Mobility Shift Assay (EMSA) with $\text{Treslin}_{\text{SHD-ext}}$ (residues 626–1053) and fluorescein-labelled dsDNA. **(c)** As in (b), but using $\text{Treslin}_{\text{ext}}$ (residues 813–1053). Protein concentrations used in (b) and (c) are serial dilution from 40 to 0 $\mu$M. **(d)** Fluorescence anisotropy experiment with $\text{Treslin}_{\text{SHD-ext}}$ and 6FAM-labelled dsDNA. **(e)** As in (d), except using $\text{Treslin}_{\text{ext}}$. **(f)** G-quadruplex motif sequence and the schematic showing the G-quadruplex secondary structure, adapted from (Maizels and Gray, 2013). **(g)** Fluorescent anisotropy experiment with $\text{Treslin}_{\text{SHD-ext}}$ and G-quadruplex (G4) or control sequence (CG4).

concentration of the protein in which roughly half of it binds to the DNA and the other half is free. The EMSA result showed that both Treslin constructs could bind to fluorescently-labelled dsDNA equally well (figure 4.14b and c). Judging from the gel shift, the $K_d$ of the $Treslin_{SHD\text{-}ext}$ for the dsDNA was estimated to be 2.5 $\mu$M.

To confirm these results and to quantify the binding affinity of the Treslin–DNA complex more accurately, fluorescence anisotropy experiments were performed using the same DNA oligomer used in EMSA experiment. Fluorescently-labelled free DNA tumbles randomly in solution. When subjected to plane polarised light, the rate of tumbling is faster than the rate of emission which results in an emitted light that is fully depolarised. Meanwhile, when labelled DNA is bound to the protein, the rate of tumbling in solution becomes slower than the rate of emission resulting in polarised light. From this phenomenon, the anisotropy measurement could be used to monitor the DNA-protein complex formation. The fluorescence anisotropy experiment result showed $K_d$ for both Treslin constructs with dsDNA were within the same order of magnitude, 2.7 $\mu$M ($\pm$0.36 $\mu$M) for $Treslin_{SHD\text{-}ext}$ and 1.95 $\mu$M ($\pm$0.41 $\mu$M) for $Treslin_{ext}$ (figure 4.14d and e).

A particular sequence that is present around the DNA replication origin consists of multiple guanine repeats that may adopt a distinguished secondary structure called G-quadruplex (G4) (Prioleau, 2017). Two to four rigid stacks of planar tetramers were formed by the guanine with negative charge present in the centre of the structure. In turn, monovalent ions such as $K^+$ or $Na^+$ helps to stabilise this structure (figure 4.14f) (Maizels and Gray, 2013). Since Treslin is involved in Cdc45 recruitment to the DNA replication origin, it would be interesting to test whether Treslin is able to recognise a DNA sequence that forms the G4. The fluorescent anisotropy experiment was repeated by using a G4 sequence. G4 has a unique circular dichroism (CD) spectra which could be used to confirm stable G4 folding. The CD confirmation was performed by Dr Joseph Maman at the Department of Biochemistry, Cambridge.

The 6FAM-labelled G4 DNA sequence (figure 4.14g) was used for the experiment with purified $Treslin_{SHD\text{-}ext}$. As a control, a sequence that is unable to form the G4 by mutating four guanine in the repeats to adenine (CG4 sequence) was also used. The steady state result from the experiments produced $K_d$ of 100 nM ($\pm$11 nM) for $Treslin_{SHD\text{-}ext}$ to the G4 sequence. In comparison, the $K_d$ of $Treslin_{SHD\text{-}ext}$ to the CG4 sequence is 2.5 $\mu$M ($\pm$0.25 $\mu$M) (figure 4.14g), which was an order of magnitude higher than the $K_d$ of Treslin with G4. Taken together, these results suggested that $Treslin_{SHD\text{-}ext}$ contains a novel DNA-binding domain. The domain is likely located in the 'extension' ($Treslin_{ext}$). It has a binding preference to the DNA sequence forming the G-quadruplex.

## 4.3 Conclusions

The Mcm2–7 hetero-hexameric DNA helicase requires incorporation of the protein co-activators, Cdc45 and GINS, for robust DNA unwinding activity. In higher eukaryotes, Treslin (the functional homologue of *S. cerevisiae*'s Sld3) is instrumental in coordinating the stepwise association of Cdc45 and GINS with the Mcm2–7 ring (Boos et al., 2011; Kumagai et al., 2010; Sansam et al., 2010; Yeeles et al., 2015). *S. cerevisiae* Sld3 is known to interact with the Mcm2–7 proteins (Deegan et al., 2016), but there is no information whether human Treslin binds to Mcm2–7 and how their putative interaction could be regulated. The principal aims of the experiments reported in this chapter were to establish whether direct binding between Treslin and Mcm2–7 exists and to provide a mechanistic description of their interaction.

The results of pulldown experiments showed that the interaction between Mcm2–7 and Treslin previously reported in *S. cerevisiae* is also present in the human system (Deegan et al., 2016). This interaction is mediated by a conserved region in vertebrate Treslin, termed $\text{Treslin}_{\text{ext}}$, that is located immediately in the C-terminal end of the Cdc45-binding domain SHD. In contrast to what was observed with the *S. cerevisiae* proteins, the interaction of Treslin with Mcm2–7 did not require DDK phosphorylation. As the Mcm2–7 sample used for the experiment was produced in HEK-293F mammalian cells, the possibility existed that it could be phosphorylated during protein expression. However, $\lambda$-phosphatase treatment of Mcm2–7 did not inhibit Treslin binding. A complete DDK phosphorylation map on Mcm2–7 would be useful in confirming whether DDK is required for Treslin–Mcm2–7 interaction in human system.

Another key finding presented in this chapter is that the integrity of the Treslin–Mcm2–7 complex is dependent on S-CDK phosphorylation. It was demonstrated that S-CDK phosphorylation disrupted the complex formation of Mcm2–7 and Treslin, in which phosphorylation of $\text{Treslin}_{\text{SHD-ext}}$ by S-CDK rendered it unable to bind to the Mcm2–7. The $\text{Treslin}_{\text{SHD-ext}}$ construct used in the experiments contains six conserved putative S-CDK sites, rather than the two known S-CDK sites on the *S. cerevisiae* Sld3. No single alanine mutation of the six S-CDK sites relieved the negative effect of S-CDK phosphorylation on the Treslin–Mcm2–7 interaction. A substantial relief of the inhibitory effect of S-CDK phosphorylation was only observed upon mutation of all of these six sites. This observation indicated a degree of functional equivalence among the six S-CDK sites. This redundancy might exist to ensure effective disruption of the Treslin–Mcm2–7 interaction at the appropriate time during CMG assembly. Thus, these biochemical data support a novel role of S-CDK in the regulation of Treslin–Mcm2–7 complex formation.

The chaperone activity of Treslin was also demonstrated in a Cdc45-Mcm2–7-Treslin reconstitution experiments. In the experiment, purified TwinStrep-tagged Mcm2–7 was bound to the StrepTactin agarose beads, before adding either Cdc45 alone, or Treslin alone, or a pre-formed Cdc45–Treslin complex. The result showed that the presence of Treslin clearly increases the association of Cdc45 to the immobilised Mcm2–7, compared with the weak binding of Cdc45 by itself. Association of Treslin and Cdc45 with the Mcm2–7 is significantly reduced by S-CDK phosphorylation, as observed with the binary Treslin–Mcm2–7 complex. These data provide support and offer a mechanistic insight on Treslin's role as a chaperone to mediate Cdc45 recruitment to Mcm2–7 during CMG assembly.

As a part of this work, the structure of human apo-Mcm2–7 complex at a 6.8 Å resolution was also reported. The structure was obtained via single-particle cryo-EM analysis. Mcm2–7 single hexamer complex adopts an open ring 'lock-washer' conformation, with a possible gate between the Mcm2 and Mcm5 and extra density in the central channel is clearly visible. This is consistent with the previous observation of MCM-SH in *S. cerevisiae* (Zhai et al., 2017a). A complex of Mcm2–7 and Treslin was also analysed by cryo-EM, but 3D reconstructions did not produce evidence of Treslin-bound Mcm2–7.

This chapter also confirmed the conserved mechanism of Treslin–TopBP1 binding across the eukaryotes (Boos et al., 2011; Kumagai et al., 2011). By using purified recombinant human proteins, the complex of Treslin and TopBP1 was shown to be dependent on S-CDK phosphorylation. The phosphorylation events on threonine 969 (T969) and serine 1001 (S1001) residues create phospho-binding sites for the BRCT0–BRCT2 repeats of TopBP1 NTD. Interestingly, a phosphomimetic mutant of $\text{Treslin}_{\text{SHD-ext}}$ (T969E + S1001D) was unable to reconstitute the interaction with TopBP1 NTD, indicating that phosphoserine and phosphothreonine play a specific role in Treslin–TopBP1 complex formation.

A novel DNA-binding activity of Treslin, which maps in the $\text{Treslin}_{\text{ext}}$ region, was also described in this chapter. EMSA and fluorescent anisotropy experiments were used to verify the DNA binding ability and to measure the binding affinity ($K_d$) of Treslin. The $K_d$ of Treslin to dsDNA was measured to be 1.95 to 3.5 $\mu$M by fluorescence anisotropy. Intriguingly, Treslin seems to have a binding preference for DNA forming a G-quadruplex (G4) structure. The $K_d$ of Treslin for G4 DNA was measured to be 100 nM by fluorescence anisotropy, at least an order of magnitude tighter than a control sequence that was unable to form G4.

This chapter describes findings that improve our understanding on how Treslin promotes DNA replication initiation, by acting as a molecular chaperone for Cdc45. The interactions between Treslin and Mcm2–7 was shown to be independent of DDK phosphorylation, unlike reported for *S. cerevisiae* Sld3, but negatively regulated by S-CDK phosphorylation. This indicates the critical role of Treslin in mediating CMG assembly is regulated by a sophisti-

cated network of phosphorylation events. Future work in cellular system will be required to validate and provide a clearer biological context for these biochemical observations.

# Chapter 5

# General discussions

## 5.1 Introduction

The general aim of this project was to understand the stepwise assembly and activation of eukaryotic replicative DNA helicase, the CMG complex. This step is highly regulated to ensure that in a single DNA replication origin, the activation of CMG helicase only happens once per cell cycle. If this step goes unregulated, it is likely to disrupt efficient DNA replication and cell division processes. Most of our knowledge about DNA replication comes from studies using model organisms, such as *S.cerevisiae* and *X. laevis*, although the details remain unclear.

This project was focused mainly on the role of Treslin in the assembly and activation of the CMG helicase. Treslin was shown to be the functional orthologue of fungal Sld3, in which they share a common Sld3/Treslin Homology Domain or SHD (Kumagai et al., 2010; Sanchez-Pulido et al., 2010; Sansam et al., 2010). One of the roles of Sld3 is to aid the recruitment of helicase co-activator, Cdc45, to the DNA-bound Mcm2–7 (Itou et al., 2014; Kamimura et al., 2001; Tanaka et al., 2011a). Furthermore, Sld3 was found to form an interaction with Mcm2–7 (Deegan et al., 2016). These observations were not yet verified using human proteins. Furthermore, human Treslin is likely to confer additional layers of complexity and regulation to its function compared to Sld3, as it is at least three times larger in size (1910 vs 668 amino acids). This project characterised Treslin interactions with other replication proteins — especially Cdc45 and Mcm2–7 — through biochemical, biophysical, and structural biology experiments.

The Cdc45–Treslin interaction was described in **chapter 3**. It was shown that human Treslin contains an additional Cdc45-binding site outside of the conserved Sld3/Treslin Homology Domain (SHD), in a region termed Treslin 'extension' ($\text{Treslin}_{\text{ext}}$). The interaction of $\text{Treslin}_{\text{ext}}$ with Cdc45 is abrogated by S-CDK phosphorylation, suggesting a mechanism

of Cdc45–Treslin regulation. Low resolution reconstruction of Cdc45–Treslin was also described in chapter 3.

Mcm2–7–Treslin complex formation was discussed in **chapter 4**. Direct interaction of human Treslin with Mcm2–7 was described for the first time, which is mediated by $Treslin_{ext}$ region. It was also shown that Mcm2–7–Treslin formation was inhibited by S-CDK phosphorylation, further highlighting the extensive role of S-CDK in regulating the assembly of CMG helicase. Furthermore, $Treslin_{ext}$ was demonstrated to possess a novel DNA-binding ability. Medium resolution reconstruction of Mcm2–7 single hexamer was also presented in chapter 4.

This final chapter aims to summarise the findings presented in this dissertation and to discuss their implications on our knowledge of eukaryotic DNA replication.

## 5.2 Summary and discussions

### 5.2.1 Cdc45–Treslin interaction

Studies in *S. cerevisiae* Sld3 demonstrated that its interaction with Cdc45 is mediated by a conserved SHD domain (Itou et al., 2014). This observation provided a molecular basis to describe Sld3-dependent Cdc45 recruitment to the DNA-bound Mcm2–7. There is no information whether human Treslin shares this Cdc45-binding activity and how this interaction could be regulated. There was no description whether a substantially larger Treslin contains additional Cdc45-binding domain outside the SHD. The structural information of Cdc45–Treslin was also unknown.

The results presented in chapter 3 verified SHD as a Cdc45-binding domain on Treslin. Purified recombinant human $Treslin_{SHD}$ is capable of interacting directly with Cdc45. The oligomeric state in solution of the Cdc45–$Treslin_{SHD}$ is in 1-to-1 ratio, as judged by SEC-MALS. The binding affinity constant ($K_d$) and the kinetic values ($K_{on}$ and $K_{off}$ of the Cdc45–$Treslin_{SHD}$ were measured by using both BLI and SPR techniques (chapter 3, table 3.1).

This project also uncovered an additional Cdc45-binding site on vertebrate Treslin. The binding site is present in $Treslin_{ext}$, next to the C-terminal end of the SHD (figure 5.1). SEC-MALS confirmed that $Treslin_{ext}$ is monomeric in solution. The binding of Cdc45–$Treslin_{ext}$ is weaker compared to Cdc45–$Treslin_{SHD}$ as shown by SEC-MALS and measurements by BLI and SPR (chapter 3, table 3.1). The presence of additional Cdc45-binding site, which does not exist in Sld3 *S. cerevisiae*, suggests an extra layer of complexity in the Cdc45–Treslin interaction of higher eukaryotes.

Treslin$_{ext}$ is an intrinsically disordered region, which contains predicted protein-binding sites and at least six potential S-CDK phosphorylation sites. It was demonstrated by pulldown experiment that S-CDK phosphorylation disrupts the Cdc45–Treslin$_{ext}$ complex formation. This is a major clue on the regulation of Cdc45–Treslin interaction. However, SEC-MALS results of Cdc45–Treslin$_{SHD\text{-}ext}$ showed that phosphorylated Cdc45–Treslin$_{SHD\text{-}ext}$ (Treslin construct containing both SHD and the 'extension' domain) formed a more homogenous complex across the gel filtration column compared to the non-phosphorylated sample — which was not in agreement with the pulldown results. The interpretation of S-CDK effect on Cdc45–Treslin$_{SHD\text{-}ext}$ oligomeric state is complicated by the non-specific interactions of Treslin$_{SHD\text{-}ext}$ and Treslin$_{ext}$ constructs to the matrix of the size-exclusion chromatography column. Treslin constructs that possess the 'extension' domain (either Treslin$_{SHD\text{-}ext}$ or Treslin$_{ext}$) have very poor chromatographic behaviour in low salt condition. High salt condition (300 mM NaCl) and pre-treatment with S-CDK improve the chromatographic behaviour. However, the high salt environment is not ideal for Cdc45–Treslin interaction studies as this complex disintegrates under such condition. Therefore, the SEC-MALS result of a 'more homogenous sample' after S-CDK phosphorylation was probably an artefact due to the non-specific interaction of the unphosphorylated Treslin$_{SHD\text{-}ext}$ to the column resin, which leads to a more likely scenario in which S-CDK actively prevents Cdc45–Treslin complex formation.

Measurements of the $K_d$ and the kinetics of the Cdc45–Treslin$_{SHD\text{-}ext}$ complex were also attempted by using both BLI and SPR to gain better better understanding on Treslin$_{ext}$ role in binding to Cdc45. The 'sticky' property of the Treslin$_{ext}$, presumably due to its highly disordered nature, caused reproducible strange phenomenon in which the analyte could not dissociate from the immobilised ligand despite screening various buffer conditions. It is possible that this was an experimental artefact caused by the biosensors-based methods, although at our current knowledge it is not possible to rule out the possibility whether it was a true biological phenomenon involving Cdc45–Treslin complex.

Structural information of Cdc45–Treslin is needed to gain deeper insight on how this complex is being formed and regulated. Unfortunately, obtaining high-resolution structures of Cdc45–Treslin proves to be a very challenging endeavour despite extensive effort in optimising samples for X-ray crystallography and single-particle cryo-EM analyses. Low-resolution 3D reconstructed maps of Cdc45–Treslin$_{SHD}$ and Cdc45–Treslin$_{SHD\text{-}ext}$ were obtained. However, in such poorly resolved maps, it is impossible to determine unambiguously the precise locations of each protein component.

Inability to obtain a higher-resolution 3D reconstruction map of the Cdc45–Treslin complex may be caused by the inherently small size of the protein complex. The size of

Cdc45–Treslin$_{SHD}$ is approximately 87.1 kDa, whereas Cdc45-Treslin$_{SHD\text{-}ext}$ is 115 kDa. For protein sample less than 150 kDa, it is challenging to obtain well-aligned particles for further processing. Currently, structures of small protein complexes that were solved with resolution better than 4 Å were using much more well-behaved samples and/or much less dynamic homo-oligomer complex (Fan et al., 2019; Khoshouei et al., 2017; Merk et al., 2016). Several optimisation methods to improve the sample quality were carried out. For sample preparation, mild cross-linking condition with or without graphene oxide were used (Goswami et al., 2018; Palovcak et al., 2018). For data collection, the Volta Phase Plate (VPP) and a small pixel size were used (Danev et al., 2014). For data processing, high-pass filtering of extracted particles to 120 Å, varying the regularisation parameter, limiting resolution E-step to 12 Å and using both RELION 3.0 and cryoSPARC 2.8.0 software packages were performed (Punjani et al., 2017; Scheres, 2016). Unfortunately, none of this step — individually or in combination — improved the quality of the data. The problem was made more complicated as the protein complexes fell apart on the grid, likely due to the air-water interface effects (Noble et al., 2018). It was also possible that the dynamic nature of the complex may contribute to the problem, as evidenced by the rapid association and dissociation rates observed during SPR and BLI measurements.

Attempts to generate full-length Treslin for biochemical and structural studies in the duration of this project were not successful due to very low expression level, either in Sf9 insect or HEK-293F mammalian cells. Co-expressing it with interacting partners — such as Cdc45, Mcm2–7, and MTBP — failed to boost the expression level. In the future, it is worth focusing on optimising full-length Treslin expression level to have enough protein sample for biochemical and structural study by cryo-EM. This could be achieved by codon-optimising the coding sequence for Treslin and MTBP for over-expression in either Sf9 or HEK-293F mammalian cells. There is evidence that codon-optimisation increases the protein yield by factoring in the availability of the tRNA which leads to improvement in the rate of elongation and accuracy of translation (Kudla et al., 2009; Plotkin and Kudla, 2011).

Once Treslin expression level is optimised to produce enough amount for structural studies, in theory it is possible to build a larger assembly of Treslin in complex with other interacting partners, such as MTBP, RecQL4, and TopBP1. This would give us better understanding of Treslin's other roles during DNA replication. It was discovered that Treslin is involved in checkpoint control during DNA damage (Guo et al., 2015; Hassan et al., 2013; Kumagai and Dunphy, 2017). These proposed experiments are dependent on the improved expression levels of codon-optimised full-length Treslin and well-behaved samples. With tremendous advancement in sample preparation for cryo-EM, it is hoped that this issue will be solved in the near future.

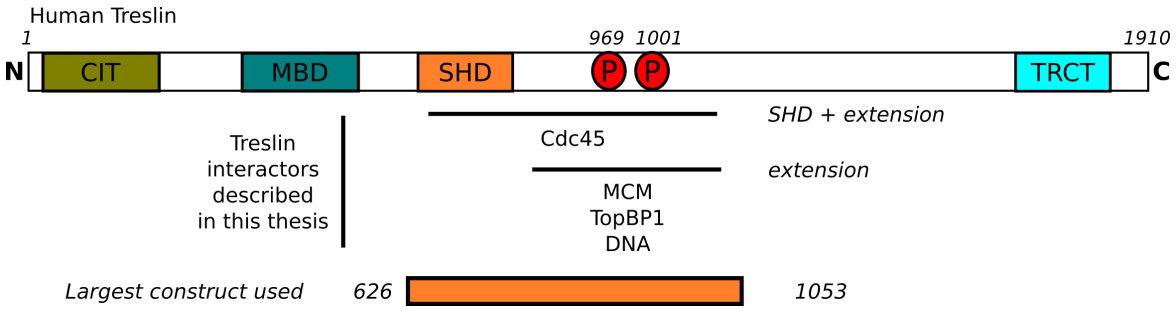



**Fig. 5.1 Mapping the protein and DNA-binding regions on human Treslin.** Summary of the regions of human Treslin that bind to other DNA replication proteins and DNA, as described in this dissertation. The largest Treslin construct used in this dissertation is indicated (Treslin residues 626–1053). MTBP and Chk1 were also found to interact with Treslin via the MBD and TRCT domains, respectively (Guo et al., 2015; Kumagai and Dunphy, 2017). CIT = Conserved in Treslin, MBD = MTBP-Binding Domain, TRCT = Treslin C-terminal, SHD = Sld3/Treslin Homology Domain. P = conserved S-CDK sites for TopBP1 binding.

In summary, the results described in this dissertation have provided several biochemical and biophysical evidences for a Cdc45–Treslin interaction in human and how it is being regulated by S-CDK phosphorylation. Furthermore, the cryo-EM analysis highlights the difficulty in obtaining a well-behaved sample for such small and dynamic protein complex.

### 5.2.2 Treslin–Mcm2–7 interaction

Sld3 was found to interact directly with Mcm2–7 in a DDK-dependent manner (Deegan et al., 2016; Heller et al., 2011; Herrera et al., 2015; Tanaka et al., 2011a). The aim of this project was to verify and to provide mechanistic insight on direct interaction between Mcm2–7 and Treslin in the human system and whether this interaction is regulated.

The results presented in this dissertation showed that Treslin is capable of forming direct interaction with the Mcm2–7 ring, as shown by pulldown experiments using purified recombinant human proteins. This interaction is mediated by $\text{Treslin}_{\text{ext}}$ (figure 5.1), which was mapped precisely to the residues 882–1053. As shown in figure 5.1, $\text{Treslin}_{\text{ext}}$ contains overlapping binding sites for Cdc45 (described in earlier section), Mcm2–7, TopBP1, and also DNA.

Intriguingly, Treslin–Mcm2–7 interaction appeared to be independent from DDK phosphorylation, in contrast with Sld3–Mcm2–7 in the fungal system. Since the Mcm2–7 was over-expressed in HEK-293F mammalian cells, it is possible that Mcm2–7 was phosphorylated *in vivo* during the expression. However, $\lambda$-phosphatase treatment failed to inhibit the Mcm2–7–Treslin formation. A complete DDK phosphorylation map on Mcm2–7 could

inform whether the DDK-independency is a true biological phenomenon in human Mcm2–7–Treslin.

A key finding was the effect of S-CDK phosphorylation on the regulation of the Mcm2–7–Treslin complex stability. When $Treslin_{ext}$ is phosphorylated, it failed to bind to the Mcm2–7. When all the six potential S-CDK sites were mutated to alanine, substantial relief of the S-CDK inhibitory effect was observed. This indicates that there is a degree of functional redundancy among the six S-CDK sites, which might help to ensure effective disruption of the Mcm2–7–Treslin complex.

The chaperone or loading factor function of Treslin for Cdc45 recruitment was demonstrated in a reconstitution experiment. There was more Cdc45 associated with the immobilised Mcm2–7 when Treslin was present, compared with weak binding of Cdc45 alone. This was a major evidence of Treslin acting as a molecular chaperone or loading factor to drive Cdc45 deposition to the Mcm2–7. Furthermore, the regulatory role of S-CDK was confirmed as association of the Cdc45–Treslin complex to the Mcm2–7 was significantly reduced by S-CDK phosphorylation.

There is no structural information on Mcm2–7–Treslin complex. The reconstituted complex of Mcm2–7–$Treslin_{SHD\text{-}ext}$ was analysed by single-particle cryo-EM. Unfortunately, the resulting 3D reconstruction map failed to provide evidence of complex formation after the vitrification process. In the future, treating the sample with chemical cross-linker could improve the sample condition. Furthermore, the chemical crosslinking experiment could map specific interacting points of Mcm2–7 with $Treslin_{SHD\text{-}ext}$. Reconstituted Mcm2–7–$Treslin_{SHD\text{-}ext}$ complex sample can be subjected to either NHS-ester cross-linker with different spacer arm length — for example DST (6.4 Å) or BS3 (11.4 Å). The samples are then subjected to mass spectrometry. The peptide coverages for both samples are then compared to sample without cross-linking treatment.

In summary, the results described in this dissertation confirmed the direct interaction between Mcm2–7–Treslin. This complex formation is likely to be DDK-independent but negatively regulated by S-CDK phosphorylation.

### 5.2.3 S-CDK regulation of CMG helicase assembly

The two minimal targets of S-CDK, to ensure normal progression of DNA replication in *S. cerevisiae*, are Sld2 and Sld3 (Tanaka et al., 2007; Zegerman and Diffley, 2007). Once phosphorylated, these two S-CDK substrates coordinate downstream events of origin firing which culminates in the formation of CMG helicase. Two highly conserved S-CDK sites on Sld3 (S600 and S622) were shown to be the binding sites of Dpb11/TopBP1 (figure 5.1),

**Table 5.1** S-CDK control on Treslin's interactions with Cdc45 and Mcm2–7

| Cell cycle phase | S-CDK level | Replication occurs? | Treslin interactions with Mcm2–7 and Cdc45? |
|---|---|---|---|
| M/G1 | Low | No | Yes |
| G1 | Low | No | Yes |
| G1/S transition | High | Yes | No |
| M | High | Yes | No |
| G2/M transition | High | No | No |

which in turn drives the recruitment of other replisome components such as GINS and Pol$\varepsilon$ (Boos et al., 2011).

One major question is the regulation for Sld3/Treslin disengagement with other replisome components. It was shown that Sld3 is only required during DNA replication initiation step, but not progression (Kanemaki and Labib, 2006). Co-immunoprecipitation assay of replication fork-associated replisome did not detect the presence of Sld3/Treslin (Gambus et al., 2006). Therefore, one of the goals of this project was to provide mechanistic understanding of how Treslin interactions with two of the CMG components are regulated.

As described in the earlier sections, S-CDK plays a major role in regulating Cdc45–Treslin and Mcm2–7–Treslin interactions. $\text{Treslin}_{\text{ext}}$ contains at least six potential S-CDK sites which regulate its interactions with Cdc45 and Mcm2–7. It needs to be noted that the two conserved S-CDK sites responsible for TopBP1 binding — threonine 969 and serine 1001 — are not directly responsible for regulating the interactions. Results from the pulldown experiments using constructs harbouring alanine mutants on T969 and S1001 did not give observable difference in binding to either Cdc45 or Mcm2–7 compared to the wild-type Treslin. Inhibitory effect of S-CDK on the Mcm2–7–Treslin binding could only be observed when using constructs harbouring alanine mutants in all six sites. This suggests a degree of functional equivalence among these sites, which might exist to ensure controlled disruption of the interactions in appropriate time. Therefore, it could be concluded that S-CDK phosphorylation holds an important role in regulating the disengagement of Treslin from both Cdc45 and Mcm2–7 prior to DNA unwinding activity.

A mechanistic model of S-CDK control of Treslin's interaction with Cdc45 and Mcm2–7 is summarised in table 5.1. S-CDK level is low during M/G1 transition and S-phase, which permits the interactions of Treslin with Cdc45 and Mcm2–7. The complex formation of Treslin–Cdc45–Mcm2–7 drives the assembly of the CMG complex on DNA replication origin. However, when S-CDK level is high during G1/S transition and S-phase, those interactions are disrupted. This S-CDK-dependent inhibition of Treslin's interactions with

Cdc45 and Mcm2–7 is crucial to prevent DNA replication origin re-firing by inhhibiting the CMG helicase formation on DNA.

Future work using cellular system may give a better understanding of Treslin's impact during cell division and growth. It would be interesting to see a cell line harbouring 6A mutant of Treslin (S820A, S838A, S923A, T969A, S1001A, and S1013A), where in theory it is unaffected by S-CDK inhibition and hence associating with both Cdc45 and Mcm2–7 longer than it should. In addition, Treslin 6A mutant should no longer be able to bind to TopBP1, which means prevention of recruitment of other replisome components such as GINS and Pol$\varepsilon$. It is likely that the replication system of this cell line would be greatly disrupted and causes cell lethality. One *in vivo* study that utilised human cells harbouring phosphomimic of Treslin in the TopBP1-binding site (T969E and S1001E) showed that DNA replication could advance but S-phase was shortened considerably (Sansam et al., 2015). It would be interesting to see whether other S-CDK sites of Treslin are responsible for regulating the normal length of S-phase to perform DNA replication.

Alternatively, the effect of Treslin 6A mutant can be determined by reconstituting DNA replication *in vitro* using purified human components, similar to the system described by Yeeles *et al* using *S. cerevisiae* proteins (Yeeles et al., 2015). It will be more straightforward to pintpoint particular step of replication that is affected by a Treslin 6A mutant. However, this would take a large effort — both in resources and time — for cloning, expressing, and purifying each of the individual human protein components responsible for DNA replication initiation.

### 5.2.4 Novel DNA-binding domain on Treslin

A DNA contamination was always observed during purification of Treslin constructs harbouring the 'extension' domain (chapter 3, figure 3.8b). This is a common problem when purifying DNA-binding DNA replication proteins. However, there is no report in the literature that Sld3/Treslin is a DNA-binding protein.

In this dissertation, a novel DNA-binding domain in $\text{Treslin}_{\text{ext}}$ was reported. The DNA-binding studies were performed using EMSAs and fluorescence anisotropy. Treslin is able to bind both single-stranded or double-stranded DNA. The $K_d$ of Treslin to dsDNA was measured by fluorescence anisotropy, which gave value of 1.95 to 2.7 $\mu$M. Intriguingly, Treslin preferentially binds to DNA sequence that forms a G-quadruplex (G4) structure. The G4 sequence contains multiple repeats of guanine that forms rigid planar stacks stabilised by monovalent ions, which is commonly found in the DNA replication origins (Prioleau, 2017). The fluorescence anisotropy experiment measured the $K_d$ of Treslin to G4 DNA to be

100 nM, which is an order of magnitude tighter compared to a control sequence that cannot adopt G4 structure ($K_d$ = 2.5 $\mu$M).

The biological relevance of Treslin binding to the G4 DNA is still unclear. It is possible that Treslin orients itself in the G4 DNA regions associated with DNA replication origin to help coordinating the assembly of the CMG complex. This may be aided by its binding partner, MTBP, which was also shown to be able to interact with G4 DNA (Kumagai and Dunphy, 2017). The importance of DNA-binding ability of Treslin can be characterised further by performing functional cell-based studies to observe whether cells harbouring Treslin's mutant for DNA binding can carry out normal DNA replication and cell division. Also, there is no report of structure G4 DNA bound to a protein yet. It will be interesting to carry out structural studies — either with Nuclear Magnetic Resonance (NMR) or X-ray crystallography — of Treslin bound to G4 DNA.

### 5.2.5 Structure of Mcm2–7 single hexamer

Our structural understanding on eukaryotic Mcm2–7 has come mostly from studies using *S. cerevisiae* or *D. melanogaster* systems (Costa et al., 2014; Li et al., 2015; Noguchi et al., 2017; Zhai et al., 2017a). However, there is no structure of human Mcm2–7 reported in the literature yet.

A medium-resolution 3D reconstructed map of human Mcm2–7 single hexamer (MCM-SH) at 6.8 Å resolution, obtained from single-particle cryo-EM analysis, was presented in this dissertation. The structure displayed each of the six subunit forming an open ring structure, similar to the one observed in other model organisms. The 3D map also showed well-defined N- and C-tier of the Mcm2–7, in which it adopts a 'lock-washer' like shape. This is in contrast with the structure of eukaryotic Mcm2–7 in the context of CMG complex, in which it is more planar (Georgescu et al., 2017; Yuan et al., 2016). There was no dsDNA observed in the central channel of the MCM-SH. This may be caused by the Mcm2–5 gate that is too narrow for dsDNA insertion and extra density in the central channel (likely to be the N-tier of Mcm5) that may actively prevent accidental insertion of ssDNA. The available *S. cerevisiae* MCM-SH could be fitted in to the 3D reconstructed map to verify the finding (Zhai et al., 2017a).

There were few particles in the electron micrographs that could be accounted for by Mcm2–7 double hexamer (MCM-DH). However, the number of the particles were too low to perform reliable 2D and 3D reconstructions. Preferred orientation also caused the 'side-view' of MCM-DH to be vastly outnumbered by the 'top-view' of MCM-DH (that is undistinguishable from MCM-SH). Structures of *S. cerevisiae* MCM-DH were already reported, in which the MCM-DH samples were prepared either by expressing endogenous

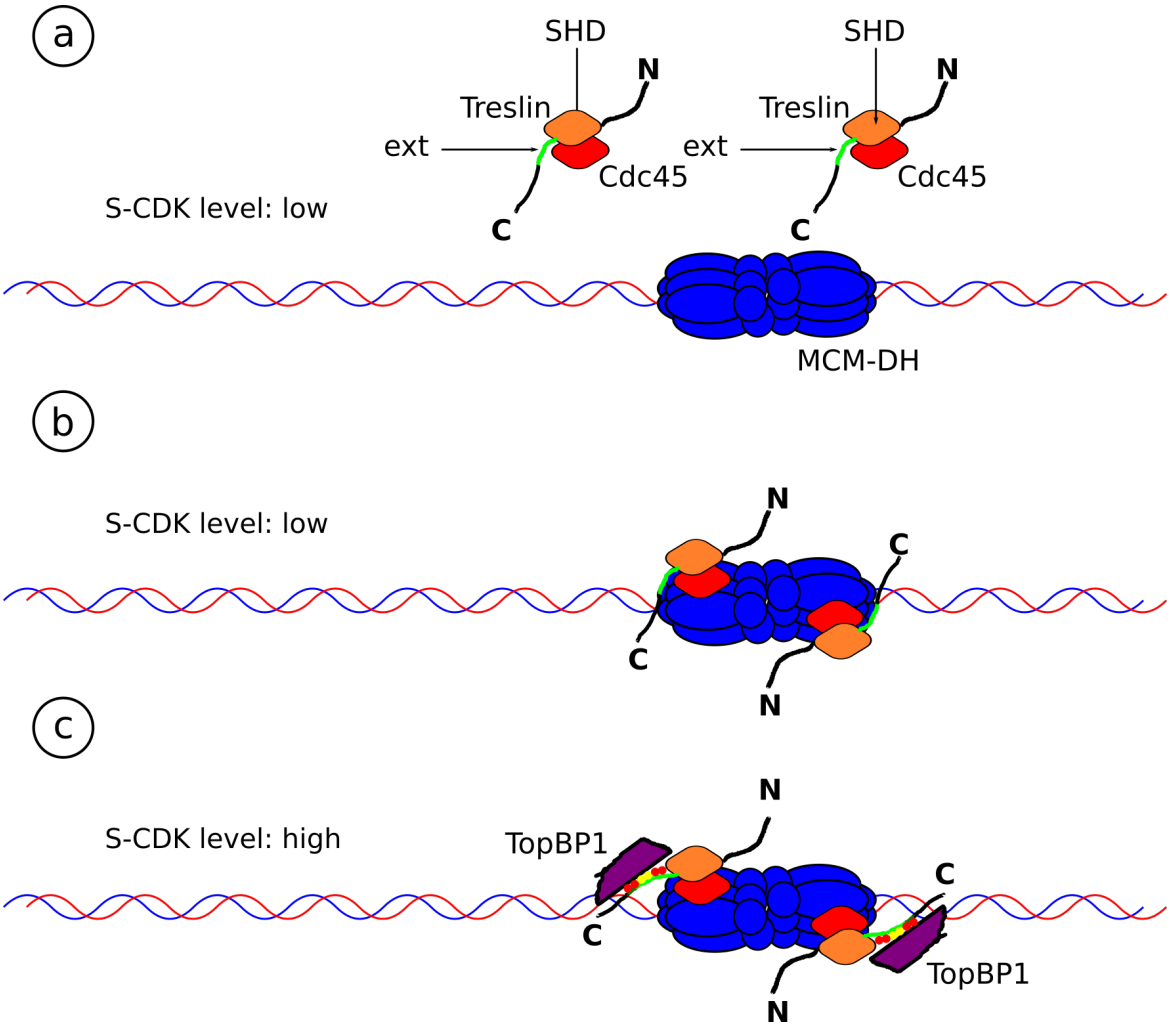



**Fig. 5.2 Proposed model for Treslin-dependent Cdc45 recruitment step to the DNA-bound Mcm2–7. (a)** Treslin interaction with Cdc45 is mediated by Sld3/Treslin Homology Domain (SHD) and 'extension' region (ext). Mcm2–7 double hexamer (MCM-DH) is assembled onto the DNA origin via ORC-dependent reaction (see chapter 1, figure 1.3). This happens during the G1 and G1/S transition phases when the S-CDK level is low. **(b)** Treslin acts as a chaperone to drive the recruitment of Cdc45 to the DNA-bound Mcm2–7. Treslin also interacts directly with the Mcm2–7 and DNA. This happens during the G1 and G1/S transition phases, when S-CDK level is low. **(c)** S-CDK phosphorylation on $\text{Treslin}_{\text{ext}}$ (represented as tiny red and yellow circles) disrupts Treslin interaction with Cdc45 and Mcm2–7. Furthermore, this allows Treslin to bind to TopBP1. Yellow circles = two conserved S-CDK sites that bind to TopBP1 (T969 and S1001), red circles = four conserved S-CDK sites that are not essential for TopBP1 binding (S820, S838, S923, S1013). N = N-terminus of Treslin, C = C-terminus of Treslin, TopBP1 = Topoisomerase II Binding Protein 1.

proteins or using ORC-dependent reaction *in vitro* (Li et al., 2015; Noguchi et al., 2017). For the experiments described in this dissertation, the Mcm2–7 sample was produced by over-expressing it in HEK-293F cells in which the minority of the sample consists of MCM-DH — confirmed by SEC-MALS — that is assembled during the expression.

Nevertheless, this is the first time that recombinant human MCM-SH structure was reported using single-particle cryo-EM analysis. Since Mcm2–7 is the core of replisome and it could be expressed well using HEK-293F cells, this provides a foundation to build a larger human replisome assembly. Once Treslin expression level is optimised, it is also possible to build a larger assembly of full-length Mcm2–7–Treslin complex for single-particle cryo-EM analysis.

## 5.3 Conclusion

The aim of this project was to understand how Treslin is involved in the process of the CMG helicase assembly in human system. The interactions of Treslin with its binding partners, such as Cdc45, Mcm2–7, TopBP1, and DNA, were characterised through a combination of biochemical, biophysical, and structural studies. The complex formations between Treslin–Mcm2–7 and Treslin–Cdc45 were also shown to be negatively regulated by S-CDK phosphorylation.

Figure 5.2 summarised a proposed model on how Cdc45 is being recruited to the Mcm2–7 by Treslin. When S-CDK level is low during the cell cycle (such as during G1/S transition), Cdc45 is able to bind to Treslin mediated by the SHD and extension (figure 5.2a). Treslin then helps to deposit Cdc45 to the DNA replication origin-bound Mcm2–7. Again, when the S-CDK activity is negligible, the complex of Treslin–Cdc45–Mcm2–7 is able to form (figure 5.2b). However, as cell cycle progresses to the S-phase and S-CDK level is rising, the interaction of Treslin–Mcm2–7 is disrupted while the Cdc45–Treslin is also weakened. In addition, as S-CDK phosphorylates at least six sites on the $\text{Treslin}_{\text{ext}}$, the complex of Treslin–TopBP1 is being formed which promotes subsequent recruitment of Pol$\varepsilon$ and GINS(Boos et al., 2011; Kumagai et al., 2010) (figure 5.2c). S-CDK regulation of Treslin's binding to its interacting partners (Cdc45 and Mcm2–7) is likely to be the mechanism on preventing re-initiation of DNA replication when it has already begun. During the replication window, S-CDK level is high which in theory inhibits Treslin's ability to recruit Cdc45 to the Mcm2–7, therefore prevents assembly of the CMG helicase when cell cycle is at the S-phase.

Since Sld3/Treslin discovery in genetic screening, its functions during DNA replication is gradually being uncovered. This dissertation provides a further biochemical understanding of Treslin's role in replicative DNA helicase assembly during DNA replication initiation.

# References


Abid Ali, F., Douglas, M. E., Locke, J., Pye, V. E., Nans, A., Diffley, J. F. X., and Costa, A. (2017). Cryo-EM structure of a licensed DNA replication origin. *Nat Commun*, 8(1):2241.

Abid Ali, F., Renault, L., Gannon, J., Gahlon, H. L., Kotecha, A., Zhou, J. C., Rueda, D., and Costa, A. (2016). Cryo-EM structures of the eukaryotic replicative helicase bound to a translocation substrate. *Nat Commun*, 7:10708.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2014). *Molecular Biology of the Cell*. Garland Publishing Inc.

Alver, R. C., Zhang, T., Josephrajan, A., Fultz, B. L., Hendrix, C. J., Das-Bradoo, S., and Bielinsky, A. K. (2014). The N-terminus of Mcm10 is important for interaction with the 9-1-1 clamp and in resistance to DNA damage. *Nucleic Acids Res.*, 42(13):8389–8404.

Aparicio, O. M., Weinstein, D. M., and Bell, S. P. (1997). Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase. *Cell*, 91(1):59–69.

Aria, V. and Yeeles, J. T. P. (2018). Mechanism of Bidirectional Leading-Strand Synthesis Establishment at Eukaryotic DNA Replication Origins. *Mol. Cell*, 73(2):199–211.

Arias, E. E. and Walter, J. C. (2007). Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Dev.*, 21(5):497–518.

Ayyagari, R., Gomes, X. V., Gordenin, D. A., and Burgers, P. M. (2003). Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 and DNA2. *J. Biol. Chem.*, 278(3):1618–1625.

Barlow, J. H. and Nussenzweig, A. (2014). Replication initiation and genome instability: a crossroads for DNA and RNA synthesis. *Cell. Mol. Life Sci.*, 71(23):4545–4559.

Bell, S. P. and Labib, K. (2016). Chromosome Duplication in Saccharomyces cerevisiae. *Genetics*, 203(3):1027–1067.

Bell, S. P. and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, 357(6374):128–134.

Bellelli, R., Belan, O., Pye, V. E., Clement, C., Maslen, S. L., Skehel, J. M., Cherepanov, P., Almouzni, G., and Boulton, S. J. (2018). POLE3-POLE4 Is a Histone H3-H4 Chaperone that Maintains Chromatin Integrity during DNA Replication. *Mol. Cell*, 72(1):112–126.

Berger, I. and Craig, A. (2011). *MultiMam Expression System User Manual*.

Betous, R., Mason, A. C., Rambo, R. P., Bansbach, C. E., Badu-Nkansah, A., Sirbu, B. M., Eichman, B. F., and Cortez, D. (2012). SMARCAL1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. *Genes Dev.*, 26(2):151–162.

Bianchi, J., Rudd, S. G., Jozwiakowski, S. K., Bailey, L. J., Soura, V., Taylor, E., Stevanovic, I., Green, A. J., Stracker, T. H., Lindsay, H. D., and Doherty, A. J. (2013). PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Mol. Cell*, 52(4):566–573.

Bicknell, L. S., Bongers, E. M., Leitch, A., Brown, S., Schoots, J., Harley, M. E., Aftimos, S., Al-Aama, J. Y., Bober, M., Brown, P. A., van Bokhoven, H., Dean, J., Edrees, A. Y., Feingold, M., Fryer, A., Hoefsloot, L. H., Kau, N., Knoers, N. V., Mackenzie, J., Opitz, J. M., Sarda, P., Ross, A., Temple, I. K., Toutain, A., Wise, C. A., Wright, M., and Jackson, A. P. (2011). Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat. Genet.*, 43(4):356–359.

Bleichert, F., Botchan, M. R., and Berger, J. M. (2017). Mechanisms for Initiating Cellular DNA Replication. *Science*, 355(6327).

Bochman, M. L., Bell, S. P., and Schwacha, A. (2008). Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Mol. Cell. Biol.*, 28(19):5865–5873.

Bochman, M. L. and Schwacha, A. (2007). Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step. *J. Biol. Chem.*, 282(46):33795–33804.

Bochman, M. L. and Schwacha, A. (2008). The Mcm2-7 complex has in vitro helicase activity. *Mol. Cell*, 31(2):287–293.

Boos, D., Sanchez-Pulido, L., Rappas, M., Pearl, L. H., Oliver, A. W., Ponting, C. P., and Diffley, J. F. (2011). Regulation of DNA replication through Sld3-Dpb11 interaction is conserved from yeast to humans. *Curr. Biol.*, 21(13):1152–1157.

Boos, D., Yekezare, M., and Diffley, J. F. (2013). Identification of a heteromeric complex that promotes DNA replication origin firing in human cells. *Science*, 340(6135):981–984.

Bowers, J. L., Randell, J. C., Chen, S., and Bell, S. P. (2004). ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Mol. Cell*, 16(6):967–978.

Branzei, D. and Foiani, M. (2006). The Rad53 signal transduction pathway: Replication fork stabilization, DNA repair, and adaptation. *Exp. Cell Res.*, 312(14):2654–2659.

Brewster, A. S. and Chen, X. S. (2010). Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. *Crit. Rev. Biochem. Mol. Biol.*, 45(3):243–256.

Bryan, T. M. (2019). Mechanisms of dna replication and repair: Insights from the study of g-quadruplexes. *Molecules*, 24(19):3439.

Bubeck, D., Reijns, M. A., Graham, S. C., Astell, K. R., Jones, E. Y., and Jackson, A. P. (2011). PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res.*, 39(9):3652–3666.

Cadoret, J. C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M. N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, 105(41):15837–15842.

Caillat, C. and Perrakis, A. (2012). Cdt1 and geminin in DNA replication initiation. *Subcell. Biochem.*, 62:71–87.

Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., and Labib, K. (2005). Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes Dev.*, 19(16):1905–1919.

Can, G., Kauerhof, A. C., Macak, D., and Zegerman, P. (2019). Helicase Subunit Cdc45 Targets the Checkpoint Kinase Rad53 to Both Replication Initiation and Elongation Complexes after Fork Stalling. *Mol. Cell*, 73(3):562–573.

Cardozo, T. and Pagano, M. (2004). The SCF ubiquitin ligase: insights into a molecular machine. *Nat. Rev. Mol. Cell Biol.*, 5(9):739–751.

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Mechali, M. (2012). New insights into replication origin characteristics in metazoans. *Cell Cycle*, 11(4):658–667.

Chattopadhyay, S. and Bielinsky, A. K. (2007). Human Mcm10 regulates the catalytic subunit of DNA polymerase-alpha and prevents DNA damage during replication. *Mol. Biol. Cell*, 18(10):4085–4095.

Chen, P., Dong, L., Hu, M., Wang, Y. Z., Xiao, X., Zhao, Z., Yan, J., Wang, P. Y., Reinberg, D., Li, M., Li, W., and Li, G. (2018). Functions of FACT in Breaking the Nucleosome and Maintaining Its Integrity at the Single-Nucleosome Level. *Mol. Cell*, 71(2):284–293.

Chen, S. and Bell, S. P. (2011). CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes Dev.*, 25(4):363–372.

Cheng, K., Xu, H., Chen, X., Wang, L., Tian, B., Zhao, Y., and Hua, Y. (2016). Structural basis for DNA 5´-end resection by RecJ. *Elife*, 5:e14294.

Cheng, Y. (2018). Single-particle cryo-EM-How did it get here and where will it go. *Science*, 361(6405):876–880.

Chesnokov, I., Remus, D., and Botchan, M. (2001). Functional analysis of mutant and wild-type Drosophila origin recognition complex. *Proc. Natl. Acad. Sci. U.S.A.*, 98(21):11997–12002.

Chilkova, O., Stenlund, P., Isoz, I., Stith, C. M., Grabowski, P., Lundstrom, E. B., Burgers, P. M., and Johansson, E. (2007). The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res.*, 35(19):6588–6597.

Cho, W. H., Kang, Y. H., An, Y. Y., Tappin, I., Hurwitz, J., and Lee, J. K. (2013). Human Tim-Tipin complex affects the biochemical properties of the replicative DNA helicase and DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.*, 110(7):2523–2527.

Cho, W. H., Lee, Y. J., Kong, S. I., Hurwitz, J., and Lee, J. K. (2006). CDC7 kinase phosphorylates serine residues adjacent to acidic amino acids in the minichromosome maintenance 2 protein. *Proc. Natl. Acad. Sci. U.S.A.*, 103(31):11521–11526.

Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J., and Blundell, T. L. (1990). X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 A resolution. *J. Mol. Biol.*, 214(1):199–222.

Cortez, D. (2019). Replication-Coupled DNA Repair. *Mol. Cell*, 74(5):866–876.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M. R., and Berger, J. M. (2011). The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat. Struct. Mol. Biol.*, 18(4):471–477.

Costa, A., Renault, L., Swuec, P., Petojevic, T., Pesavento, J. J., Ilves, I., MacLellan-Gibson, K., Fleck, R. A., Botchan, M. R., and Berger, J. M. (2014). DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *Elife*, 3:e03273.

Costanzo, V., Robertson, K., Bibikova, M., Kim, E., Grieco, D., Gottesman, M., Carroll, D., and Gautier, J. (2001). Mre11 protein complex prevents double-strand break accumulation during chromosomal DNA replication. *Mol. Cell*, 8(1):137–147.

Coster, G., Frigola, J., Beuron, F., Morris, E. P., and Diffley, J. F. (2014). Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. *Mol. Cell*, 55(5):666–677.

Danev, R. and Baumeister, W. (2017). Expanding the boundaries of cryo-EM with phase plates. *Curr. Opin. Struct. Biol.*, 46:87–94.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 111(44):15635–15640.

Das-Bradoo, S., Ricke, R. M., and Bielinsky, A. K. (2006). Interaction between PCNA and diubiquitinated Mcm10 is essential for cell growth in budding yeast. *Mol. Cell. Biol.*, 26(13):4806–4817.

Davey, M. J., Indiani, C., and O'Donnell, M. (2003). Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. *J. Biol. Chem.*, 278(7):4491–4499.

De Koning, L., Corpet, A., Haber, J. E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. *Nat. Struct. Mol. Biol.*, 14(11):997–1007.

De Piccoli, G., Katou, Y., Itoh, T., Nakato, R., Shirahige, K., and Labib, K. (2012). Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases. *Mol. Cell*, 45(5):696–704.

Deegan, T. D., Baxter, J., Ortiz Bazan, M. A., Yeeles, J. T. P., and Labib, K. P. M. (2019). Pif1-Family Helicases Support Fork Convergence during DNA Replication Termination in Eukaryotes. *Mol. Cell*, 74(2):231–244.

Deegan, T. D., Yeeles, J. T., and Diffley, J. F. (2016). Phosphopeptide binding by Sld3 links Dbf4-dependent kinase to MCM replicative helicase activation. *EMBO J.*, 35(9):961–973.

DePamphilis, M. L., Blow, J. J., Ghosh, S., Saha, T., Noguchi, K., and Vassilev, A. (2006). Regulating the licensing of DNA replication origins in metazoa. *Curr. Opin. Cell Biol.*, 18(3):231–239.

Dewar, J. M., Budzowska, M., and Walter, J. C. (2015). The mechanism of DNA replication termination in vertebrates. *Nature*, 525(7569):345–350.

Diffley, J. F., Cocker, J. H., Dowell, S. J., and Rowley, A. (1994). Two steps in the assembly of complexes at yeast replication origins in vivo. *Cell*, 78(2):303–316.

Dosztanyi, Z., Meszaros, B., and Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, 25(20):2745–2746.

Douglas, M. E., Ali, F. A., Costa, A., and Diffley, J. F. X. (2018). The mechanism of eukaryotic CMG helicase activation. *Nature*, 555(7695):265–268.

Driscoll, R., Hudson, A., and Jackson, S. P. (2007). Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. *Science*, 315(5812):649–652.

Drury, L. S., Perkins, G., and Diffley, J. F. (2000). The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr. Biol.*, 10(5):231–240.

Dua, R., Levy, D. L., and Campbell, J. L. (1999). Analysis of the essential functions of the C-terminal protein/protein interaction domain of Saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J. Biol. Chem.*, 274(32):22283–22288.

Dungrawala, H., Rose, K. L., Bhat, K. P., Mohni, K. N., Glick, G. G., Couch, F. B., and Cortez, D. (2015). The Replication Checkpoint Prevents Two Types of Fork Collapse without Regulating Replisome Stability. *Mol. Cell*, 59(6):998–1010.

Dutta, A. and Bell, S. P. (1997). Initiation of DNA replication in eukaryotic cells. *Annu. Rev. Cell Dev. Biol.*, 13:293–332.

Duzdevich, D., Warner, M. D., Ticau, S., Ivica, N. A., Bell, S. P., and Greene, E. C. (2015). The dynamics of eukaryotic replication initiation: origin specificity, licensing, and firing at the single-molecule level. *Mol. Cell*, 58(3):483–494.

Errico, A., Cosentino, C., Rivera, T., Losada, A., Schwob, E., Hunt, T., and Costanzo, V. (2009). Tipin/Tim1/And1 protein complex promotes Pol alpha chromatin binding and sister chromatid cohesion. *EMBO J.*, 28(23):3681–3692.

Errico, A. and Costanzo, V. (2012). Mechanisms of replication fork protection: a safeguard for genome stability. *Crit. Rev. Biochem. Mol. Biol.*, 47(3):222–235.

Errico, A., Costanzo, V., and Hunt, T. (2007). Tipin is required for stalled replication forks to resume DNA replication after removal of aphidicolin in Xenopus egg extracts. *Proc. Natl. Acad. Sci. U.S.A.*, 104(38):14929–14934.

Errico, A., Deshmukh, K., Tanaka, Y., Pozniakovsky, A., and Hunt, T. (2010). Identification of substrates for cyclin dependent kinases. *Adv. Enzyme Regul.*, 50(1):375–399.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, 106(48):20240–20245.

Evrin, C., Fernandez-Cid, A., Riera, A., Zech, J., Clarke, P., Herrera, M. C., Tognetti, S., Lurz, R., and Speck, C. (2014). The ORC/Cdc6/MCM2-7 complex facilitates MCM2-7 dimerization during prereplicative complex formation. *Nucleic Acids Res.*, 42(4):2257–2269.

Evrin, C., Maman, J. D., Diamante, A., Pellegrini, L., and Labib, K. (2018). Histone H2A-H2B binding by Pol $\alpha$ in the eukaryotic replisome contributes to the maintenance of repressive chromatin. *EMBO J.*, 37(19).

Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J. C., Zhao, L., Peng, H. L., Lei, J., and Wang, H. W. (2019). Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat Commun*, 10(1):2386.

Ferguson, B. M. and Fangman, W. L. (1992). A position effect on the time of replication origin activation in yeast. *Cell*, 68(2):333–339.

Fernandez-Cid, A., Riera, A., Tognetti, S., Herrera, M. C., Samel, S., Evrin, C., Winkler, C., Gardenal, E., Uhle, S., and Speck, C. (2013). An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Mol. Cell*, 50(4):577–588.

Foltman, M., Evrin, C., De Piccoli, G., Jones, R. C., Edmondson, R. D., Katou, Y., Nakato, R., Shirahige, K., and Labib, K. (2013). Eukaryotic replisome components cooperate to process histones during chromosome replication. *Cell Rep*, 3(3):892–904.

Francis, L. I., Randell, J. C., Takara, T. J., Uchima, L., and Bell, S. P. (2009). Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. *Genes Dev.*, 23(5):643–654.

Frigola, J., Remus, D., Mehanna, A., and Diffley, J. F. (2013). ATPase-dependent quality control of DNA replication origin licensing. *Nature*, 495(7441):339–343.

Fu, Y. V., Yardimci, H., Long, D. T., Ho, T. V., Guainazzi, A., Bermudez, V. P., Hurwitz, J., van Oijen, A., Scharer, O. D., and Walter, J. C. (2011). Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell*, 146(6):931–941.

Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R. D., and Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat. Cell Biol.*, 8(4):358–366.

Gambus, A., Priego Moreno, S., Jones, R. M., and Poovathumkadavil, D. (2018). Mitotic replisome disassembly in vertebrates. *bioRxiv*.

Gambus, A., van Deursen, F., Polychronopoulos, D., Foltman, M., Jones, R. C., Edmondson, R. D., Calzada, A., and Labib, K. (2009). A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *EMBO J.*, 28(19):2992–3004.

Garcia-Gomez, S., Reyes, A., Martinez-Jimenez, M. I., Chocron, E. S., Mouron, S., Terrados, G., Powell, C., Salido, E., Mendez, J., Holt, I. J., and Blanco, L. (2013). PrimPol, an archaic primase/polymerase operating in human cells. *Mol. Cell*, 52(4):541–553.

Gary, R., Park, M. S., Nolan, J. P., Cornelius, H. L., Kozyreva, O. G., Tran, H. T., Lobachev, K. S., Resnick, M. A., and Gordenin, D. A. (1999). A novel role in DNA metabolism for the binding of Fen1/Rad27 to PCNA and implications for genetic risk. *Mol. Cell. Biol.*, 19(8):5373–5382.

Georgescu, R. E., Langston, L., Yao, N. Y., Yurieva, O., Zhang, D., Finkelstein, J., Agarwal, T., and O'Donnell, M. E. (2014). Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.*, 21(8):664–670.

Georgescu, R. E., Yuan, Z., Bai, L., de Luna Almeida Santos, R., Sun, J., Zhang, D., Yurieva, O., Li, H., and O'Donnell, M. E. (2017). Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc. Natl. Acad. Sci. U.S.A.*, 114(5):E697–E706.

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741.

Gibson, S. I., Surosky, R. T., and Tye, B. K. (1990). The phenotype of the minichromosome maintenance mutant mcm3 is characteristic of mutants defective in DNA replication. *Mol. Cell. Biol.*, 10(11):5707–5720.

Giordano-Coltart, J., Ying, C. Y., Gautier, J., and Hurwitz, J. (2005). Studies of the properties of human origin recognition complex and its Walker A motif mutants. *Proc. Natl. Acad. Sci. U.S.A.*, 102(1):69–74.

Goswami, P., Abid Ali, F., Douglas, M. E., Locke, J., Purkiss, A., Janska, A., Eickhoff, P., Early, A., Nans, A., Cheung, A. M. C., Diffley, J. F. X., and Costa, A. (2018). Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. *Nat Commun*, 9(1):5061.

Guo, C., Kumagai, A., Schlacher, K., Shevchenko, A., Shevchenko, A., and Dunphy, W. G. (2015). Interaction of Chk1 with Treslin negatively regulates the initiation of chromosomal DNA replication. *Mol. Cell*, 57(3):492–505.

Hanzlikova, H., Kalasova, I., Demin, A. A., Pennicott, L. E., Cihlarova, Z., and Caldecott, K. W. (2018). The Importance of Poly(ADP-Ribose) Polymerase as a Sensor of Unligated Okazaki Fragments during DNA Replication. *Mol. Cell*, 71(2):319–331.

Harding, S. E. and Jumel, K. (1998). Light scattering. *Curr Protoc Protein Sci*, 11(1):7–8.

Hardy, C. F., Dryga, O., Seematter, S., Pahl, P. M., and Sclafani, R. A. (1997). mcm5/cdc46-bob1 bypasses the requirement for the S phase activator Cdc7p. *Proc. Natl. Acad. Sci. U.S.A.*, 94(7):3151–3155.

Hartwell, L. (1992). Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell*, 71(4):543–546.

Hashimoto, Y., Ray Chaudhuri, A., Lopes, M., and Costanzo, V. (2010). Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nat. Struct. Mol. Biol.*, 17(11):1305–1311.

Hassan, B. H., Lindsey-Boltz, L. A., Kemp, M. G., and Sancar, A. (2013). Direct role for the replication protein Treslin (Ticrr) in the ATR kinase-mediated checkpoint response. *J. Biol. Chem.*, 288(26):18903–18910.

Hawkins, M., Retkute, R., Muller, C. A., Saner, N., Tanaka, T. U., de Moura, A. P., and Nieduszynski, C. A. (2013). High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep*, 5(4):1132–1141.

Heller, R. C., Kang, S., Lam, W. M., Chen, S., Chan, C. S., and Bell, S. P. (2011). Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. *Cell*, 146(1):80–91.

Hennessy, K. M., Lee, A., Chen, E., and Botstein, D. (1991). A group of interacting yeast DNA replication genes. *Genes Dev.*, 5(6):958–969.

Herrera, M. C., Tognetti, S., Riera, A., Zech, J., Clarke, P., Fernandez-Cid, A., and Speck, C. (2015). A reconstituted system reveals how activating and inhibitory interactions control DDK dependent assembly of the eukaryotic replicative helicase. *Nucleic Acids Res.*, 43(21):10238–10250.

Hills, S. A. and Diffley, J. F. (2014). DNA replication and oncogene-induced replicative stress. *Curr. Biol.*, 24(10):R435–444.

Hodgson, B., Calzada, A., and Labib, K. (2007). Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. *Mol. Biol. Cell*, 18(10):3894–3902.

Hogg, M. and Johansson, E. (2012). DNA polymerase $\varepsilon$. *Subcell. Biochem.*, 62:237–257.

Hombauer, H., Campbell, C. S., Smith, C. E., Desai, A., and Kolodner, R. D. (2011). Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell*, 147(5):1040–1053.

Hopwood, B. and Dalton, S. (1996). Cdc45p assembles into a complex with Cdc46p/Mcm5p, is required for minichromosome maintenance, and is essential for chromosomal DNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, 93(22):12309–12314.

Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., and Waga, S. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.*, 288(42):30161–30171.

Huang, H., Stromme, C. B., Saredi, G., Hodl, M., Strandsby, A., Gonzalez-Aguilera, C., Chen, S., Groth, A., and Patel, D. J. (2015). A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks. *Nat. Struct. Mol. Biol.*, 22(8):618–626.

Hughes, S., Elustondo, F., Di Fonzo, A., Leroux, F. G., Wong, A. C., Snijders, A. P., Matthews, S. J., and Cherepanov, P. (2012). Crystal structure of human CDC7 kinase in complex with its activator DBF4. *Nat. Struct. Mol. Biol.*, 19(11):1101–1107.

Ilves, I., Petojevic, T., Pesavento, J. J., and Botchan, M. R. (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol. Cell*, 37(2):247–258.

Itou, H., Muramatsu, S., Shirakihara, Y., and Araki, H. (2014). Crystal structure of the homology domain of the eukaryotic DNA replication proteins Sld3/Treslin. *Structure*, 22(9):1341–1347.

Itou, H., Shirakihara, Y., and Araki, H. (2015). The quaternary structure of the eukaryotic DNA replication proteins Sld7 and Sld3. *Acta Crystallogr. D Biol. Crystallogr.*, 71(Pt 8):1649–1656.

Jackson, A. P., Laskey, R. A., and Coleman, N. (2014). Replication proteins and human disease. *Cold Spring Harb Perspect Biol*, 6(1).

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the regulation of dna replication in bacteria. In *Cold Spring Harbor symposia on quantitative biology*, volume 28, pages 329–348. Cold Spring Harbor Laboratory Press.

Johansson, E. and Dixon, N. (2013). Replicative DNA polymerases. *Cold Spring Harb Perspect Biol*, 5(6).

Kalfalah, F., Seggewiss, S., Walter, R., Tigges, J., Moreno-Villanueva, M., Burkle, A., Ohse, S., Busch, H., Boerries, M., Hildebrandt, B., Royer-Pokora, B., and Boege, F. (2015). Structural chromosome abnormalities, increased DNA strand breaks and DNA strand break repair deficiency in dermal fibroblasts from old female human donors. *Aging (Albany NY)*, 7(2):110–122.

Kamimura, Y., Tak, Y. S., Sugino, A., and Araki, H. (2001). Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in Saccharomyces cerevisiae. *EMBO J.*, 20(8):2097–2107.

Kanemaki, M. and Labib, K. (2006). Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *EMBO J.*, 25(8):1753–1763.

Katan-Khaykovich, Y. and Struhl, K. (2011). Splitting of H3-H4 tetramers at transcriptionally active genes undergoing dynamic histone exchange. *Proc. Natl. Acad. Sci. U.S.A.*, 108(4):1296–1301.

Kawakami, H., Ohashi, E., Kanamoto, S., Tsurimoto, T., and Katayama, T. (2015). Specific binding of eukaryotic ORC to DNA replication origins depends on highly conserved basic residues. *Sci Rep*, 5:14929.

Kesti, T., Flick, K., Keranen, S., Syvaoja, J. E., and Wittenberg, C. (1999). DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol. Cell*, 3(5):679–685.

Khoshouei, M., Radjainia, M., Baumeister, W., and Danev, R. (2017). Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat Commun*, 8:16099.

Kilkenny, M. L., Simon, A. C., Mainwaring, J., Wirthensohn, D., Holzer, S., and Pellegrini, L. (2017). The human CTF4-orthologue AND-1 interacts with DNA polymerase$\alpha$/primase via its unique C-terminal HMG box. *Open Biol*, 7(11).

Kim, S., Dallmann, H. G., McHenry, C. S., and Marians, K. J. (1996). Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell*, 84(4):643–650.

Kleczkowska, H. E., Marra, G., Lettieri, T., and Jiricny, J. (2001). hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev.*, 15(6):724–736.

Klemm, R. D., Austin, R. J., and Bell, S. P. (1997). Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. *Cell*, 88(4):493–502.

Klinge, S., Hirst, J., Maman, J. D., Krude, T., and Pellegrini, L. (2007). An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis. *Nat. Struct. Mol. Biol.*, 14(9):875–877.

Knott, S. R., Viggiani, C. J., Tavare, S., and Aparicio, O. M. (2009). Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in Saccharomyces cerevisiae. *Genes Dev.*, 23(9):1077–1090.

Kose, H. B., Larsen, N. B., Duxin, J. P., and Yardimci, H. (2019). Dynamics of the Eukaryotic Replicative Helicase at Lagging-Strand Protein Barriers Support the Steric Exclusion Model. *Cell Rep*, 26(8):2113–2125.

Krastanova, I., Sannino, V., Amenitsch, H., Gileadi, O., Pisani, F. M., and Onesti, S. (2012). Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. *J. Biol. Chem.*, 287(6):4121–4128.

Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *Science*, 324(5924):255–258.

Kumagai, A. and Dunphy, W. G. (2017). MTBP, the partner of Treslin, contains a novel DNA-binding domain that is essential for proper initiation of DNA replication. *Mol. Biol. Cell*, 28(22):2998–3012.

Kumagai, A., Shevchenko, A., Shevchenko, A., and Dunphy, W. G. (2010). Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. *Cell*, 140(3):349–359.

Kumagai, A., Shevchenko, A., Shevchenko, A., and Dunphy, W. G. (2011). Direct regulation of Treslin by cyclin-dependent kinase is essential for the onset of DNA replication. *J. Cell Biol.*, 193(6):995–1007.

Kumaraswamy, S. and Tobias, R. (2015). Label-free kinetic analysis of an antibody-antigen interaction using biolayer interferometry. *Methods Mol. Biol.*, 1278:165–182.

Labib, K. (2010). How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes Dev.*, 24(12):1208–1219.

Labib, K., Kearsey, S. E., and Diffley, J. F. (2001). MCM2-7 proteins are essential components of prereplicative complexes that accumulate cooperatively in the nucleus during G1-phase and are required to establish, but not maintain, the S-phase checkpoint. *Mol. Biol. Cell*, 12(11):3658–3667.

Labib, K., Tercero, J. A., and Diffley, J. F. (2000). Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science*, 288(5471):1643–1647.

Langston, L. D., Zhang, D., Yurieva, O., Georgescu, R. E., Finkelstein, J., Yao, N. Y., Indiani, C., and O'Donnell, M. E. (2014). CMG helicase and DNA polymerase $\varepsilon$ form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, 111(43):15390–15395.

Larsen, N. B., Gao, A. O., Sparks, J. L., Gallina, I., Wu, R. A., Mann, M., Raschle, M., Walter, J. C., and Duxin, J. P. (2019). Replication-Coupled DNA-Protein Crosslink Repair by SPRTN and the Proteasome in Xenopus Egg Extracts. *Mol. Cell*, 73(3):574–588.

Li, H. and Stillman, B. (2012). The origin recognition complex: a biochemical and structural view. *Subcell. Biochem.*, 62:37–58.

Li, N., Lam, W. H., Zhai, Y., Cheng, J., Cheng, E., Zhao, Y., Gao, N., and Tye, B. K. (2018). Structure of the origin recognition complex bound to DNA replication origin. *Nature*, 559(7713):217–222.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B. K., and Gao, N. (2015). Structure of the eukaryotic MCM complex at 3.8 Å. *Nature*, 524(7564):186–191.

Ling, C., Huang, J., Yan, Z., Li, Y., Ohzeki, M., Ishiai, M., Xu, D., Takata, M., Seidman, M., and Wang, W. (2016). Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discov*, 2:16047.

Liu, Y., Kao, H. I., and Bambara, R. A. (2004). Flap endonuclease 1: a central component of DNA metabolism. *Annu. Rev. Biochem.*, 73:589–615.

Lopez-Mosqueda, J., Maas, N. L., Jonsson, Z. O., Defazio-Eli, L. G., Wohlschlegel, J., and Toczyski, D. P. (2010). Damage-induced phosphorylation of Sld3 is important to block late origin firing. *Nature*, 467(7314):479–483.

Lucchini, R. and Sogo, J. M. (1995). Replication of transcriptionally active chromatin. *Nature*, 374(6519):276–280.

Ma, X., Stead, B. E., Rezvanpour, A., and Davey, M. J. (2010). The effects of oligomerization on Saccharomyces cerevisiae Mcm4/6/7 function. *BMC Biochem.*, 11:37.

Maine, G. T., Sinha, P., and Tye, B. K. (1984). Mutants of S. cerevisiae defective in the maintenance of minichromosomes. *Genetics*, 106(3):365–385.

Maizels, N. and Gray, L. T. (2013). The G4 genome. *PLoS Genet.*, 9(4):e1003468.

Mantiero, D., Mackenzie, A., Donaldson, A., and Zegerman, P. (2011). Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J.*, 30(23):4805–4814.

Maric, M., Maculins, T., De Piccoli, G., and Labib, K. (2014). Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science*, 346(6208):1253596.

Masai, H., Taniyama, C., Ogino, K., Matsui, E., Kakusho, N., Matsumoto, S., Kim, J. M., Ishii, A., Tanaka, T., Kobayashi, T., Tamai, K., Ohtani, K., and Arai, K. (2006). Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin. *J. Biol. Chem.*, 281(51):39249–39261.

Masuda, T., Mimura, S., and Takisawa, H. (2003). CDK- and Cdc45-dependent priming of the MCM complex on chromatin during S-phase in Xenopus egg extracts: possible activation of MCM helicase by association with Cdc45. *Genes Cells*, 8(2):145–161.

Masumoto, H., Hawke, D., Kobayashi, R., and Verreault, A. (2005). A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature*, 436(7048):294–298.

Mayle, R., Langston, L., Molloy, K. R., Zhang, D., Chait, B. T., and O'Donnell, M. E. (2019). Mcm10 has potent strand-annealing activity and limits translocase-mediated fork regression. *Proc. Natl. Acad. Sci. U.S.A.*, 116(3):798–803.

McElhinny, N. S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M., and Kunkel, T. A. (2008). Division of labor at the eukaryotic replication fork. *Mol. Cell*, 30(2):137–144.

McGarry, T. J. and Kirschner, M. W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell*, 93(6):1043–1053.

McMullan, G., Faruqi, A. R., and Henderson, R. (2016). Direct Electron Detectors. *Meth. Enzymol.*, 579:1–17.

Mechali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.*, 11(10):728–738.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J. L. S., and Subramaniam, S. (2016). Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell*, 165(7):1698–1707.

Meszaros, B., Simon, I., and Dosztanyi, Z. (2009). Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.*, 5(5):e1000376.

Mimura, S., Masuda, T., Matsui, T., and Takisawa, H. (2000). Central role for cdc45 in establishing an initiation complex of DNA replication in Xenopus egg extracts. *Genes Cells*, 5(6):439–452.

Mimura, S., Seki, T., Tanaka, S., and Diffley, J. F. (2004). Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. *Nature*, 431(7012):1118–1123.

Mizushima, T., Takahashi, N., and Stillman, B. (2000). Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. *Genes Dev.*, 14(13):1631–1641.

Mol, N. J. and Fischer, M. J. (2010). *Surface plasmon resonance: methods and protocols*. Springer.

Montagnoli, A., Valsasina, B., Brotherton, D., Troiani, S., Rainoldi, S., Tenca, P., Molinari, A., and Santocanale, C. (2006). Identification of Mcm2 phosphorylation sites by S-phase-regulating kinases. *J. Biol. Chem.*, 281(15):10281–10290.

Morohashi, H., Maculins, T., and Labib, K. (2009). The amino-terminal TPR domain of Dia2 tethers SCF(Dia2) to the replisome progression complex. *Curr. Biol.*, 19(22):1943–1949.

Moyer, S. E., Lewis, P. W., and Botchan, M. R. (2006). Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci. U.S.A.*, 103(27):10236–10241.

Muramatsu, S., Hirai, K., Tak, Y. S., Kamimura, Y., and Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol $\varepsilon$, and GINS in budding yeast. *Genes Dev.*, 24(6):602–612.

Muzi-Falconi, M., Giannattasio, M., Foiani, M., and Plevani, P. (2003). The DNA polymerase alpha-primase complex: multiple functions and interactions. *ScientificWorldJournal*, 3:21–33.

Nakajima, R. and Masukata, H. (2002). SpSld3 is required for loading and maintenance of SpCdc45 on chromatin in DNA replication in fission yeast. *Mol. Biol. Cell*, 13(5):1462–1472.

Naylor, M. L., Li, J. M., Osborn, A. J., and Elledge, S. J. (2009). Mrc1 phosphorylation in response to DNA replication stress is required for Mec1 accumulation at the stalled fork. *Proc. Natl. Acad. Sci. U.S.A.*, 106(31):12765–12770.

Nguyen, V. Q., Co, C., and Li, J. J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature*, 411(6841):1068–1073.

Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S., and Carragher, B. (2018). Reducing effects of particle adsorption to the air-water interface in cryo-EM. *Nat. Methods*, 15(10):793–795.

Nogales, E. and Scheres, S. H. (2015). Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell*, 58(4):677–689.

Noguchi, Y., Yuan, Z., Bai, L., Schneider, S., Zhao, G., Stillman, B., Speck, C., and Li, H. (2017). Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-strand DNA extrusion model. *Proc. Natl. Acad. Sci. U.S.A.*, 114(45):E9529–E9538.

Nyberg, K. A., Michelson, R. J., Putnam, C. W., and Weinert, T. A. (2002). Toward maintaining the genome: DNA damage and replication checkpoints. *Annu. Rev. Genet.*, 36:617–656.

Osborn, A. J. and Elledge, S. J. (2003). Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes Dev.*, 17(14):1755–1767.

Oyama, T., Ishino, S., Shirai, T., Yamagami, T., Nagata, M., Ogino, H., Kusunoki, M., and Ishino, Y. (2016). Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Res.*, 44(19):9505–9517.

Pacek, M., Tutter, A. V., Kubota, Y., Takisawa, H., and Walter, J. C. (2006). Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol. Cell*, 21(4):581–587.

Pacek, M. and Walter, J. C. (2004). A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. *EMBO J.*, 23(18):3667–3676.

Palovcak, E., Wang, F., Zheng, S. Q., Yu, Z., Li, S., Betegon, M., Bulkley, D., Agard, D. A., and Cheng, Y. (2018). A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *J. Struct. Biol.*, 204(1):80–84.

Pan, X., Ye, P., Yuan, D. S., Wang, X., Bader, J. S., and Boeke, J. D. (2006). A DNA integrity network in the yeast Saccharomyces cerevisiae. *Cell*, 124(5):1069–1081.

Park, J. H., Bang, S. W., Kim, S. H., and Hwang, D. S. (2008). Knockdown of human MCM10 activates G2 checkpoint pathway. *Biochem. Biophys. Res. Commun.*, 365(3):490–495.

Pavletich, N. P. (1999). Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J. Mol. Biol.*, 287(5):821–828.

Pellegrini, L. (2012). The Pol $\alpha$-primase complex. *Subcell. Biochem.*, 62:157–169.

Pellegrini, L. (2017). Structural insights into Cdc45 function: was there a nuclease at the heart of the ancestral replisome? *Biophys. Chem.*, 225:10–14.

Peranen, J., Rikkonen, M., Hyvonen, M., and Kaariainen, L. (1996). T7 vectors with modified T7lac promoter for expression of proteins in Escherichia coli. *Anal. Biochem.*, 236(2):371–373.

Perera, R. L., Torella, R., Klinge, S., Kilkenny, M. L., Maman, J. D., and Pellegrini, L. (2013). Mechanism for priming DNA synthesis by yeast DNA polymerase $\alpha$. *Elife*, 2:e00482.

Petojevic, T., Pesavento, J. J., Costa, A., Liang, J., Wang, Z., Berger, J. M., and Botchan, M. R. (2015). Cdc45 (cell division cycle protein 45) guards the gate of the Eukaryote Replisome helicase stabilizing leading strand engagement. *Proc. Natl. Acad. Sci. U.S.A.*, 112(3):E249–258.

Plotkin, J. B. and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, 12(1):32–42.

Primorac, I. and Musacchio, A. (2013). Panta rhei: the APC/C at steady state. *J. Cell Biol.*, 201(2):177–189.

Prioleau, M. N. (2017). G-Quadruplexes and DNA Replication Origins. *Adv. Exp. Med. Biol.*, 1042:273–286.

Prorok, P., Artufel, M., Aze, A., Coulombe, P., Peiffer, I., Lacroix, L., Guedin, A., Mergny, J. L., Damaschke, J., Schepers, A., Ballester, B., and Mechali, M. (2019). Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat Commun*, 10(1):3274.

Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, 14(3):290–296.

Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E., and Kunkel, T. A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, 317(5834):127–130.

Pursell, Z. F. and Kunkel, T. A. (2008). DNA polymerase epsilon: a polymerase of unusual size (and complexity). *Prog. Nucleic Acid Res. Mol. Biol.*, 82:101–145.

Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D. J., Davis, R. W., Brewer, B. J., and Fangman, W. L. (2001). Replication dynamics of the yeast genome. *Science*, 294(5540):115–121.

Randell, J. C., Bowers, J. L., Rodriguez, H. K., and Bell, S. P. (2006). Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol. Cell*, 21(1):29–39.

Remus, D., Beall, E. L., and Botchan, M. R. (2004). DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *EMBO J.*, 23(4):897–907.

Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P., and Diffley, J. F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell*, 139(4):719–730.

Reyes-Lamothe, R., Sherratt, D. J., and Leake, M. C. (2010). Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science*, 328(5977):498–501.

Richet, N., Liu, D., Legrand, P., Velours, C., Corpet, A., Gaubert, A., Bakail, M., Moal-Raisin, G., Guerois, R., Compper, C., Besle, A., Guichard, B., Almouzni, G., and Ochsenbein, F. (2015). Structural insight into how the human helicase subunit MCM2 may act as a histone chaperone together with ASF1 at the replication fork. *Nucleic Acids Res.*, 43(3):1905–1917.

Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S. M., Huguenin-Dezot, N., Muqit, M. M., Fry, A. M., Bayliss, R., and Chin, J. W. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.*, 11(7):496–503.

Rohleder, F., Huang, J., Xue, Y., Kuper, J., Round, A., Seidman, M., Wang, W., and Kisker, C. (2016). FANCM interacts with PCNA to promote replication traverse of DNA interstrand crosslinks. *Nucleic Acids Res.*, 44(7):3219–3232.

Rouse, J. and Jackson, S. P. (2002). Interfaces between the detection, signaling, and repair of DNA damage. *Science*, 297(5581):547–551.

Rzechorzek, N. J., Hardwick, S. W., Jatikusumo, V. A., Chirgadze, D. Y., and Pellegrini, L. (2020). CryoEM structures of human CMG-ATPγS-DNA and CMG-AND-1 complexes. *bioRxiv*, page 2020.01.22.914192.

Sakabe, K. and Okazaki, R. (1966). A unique property of the replicating region of chromosomal DNA. *Biochim. Biophys. Acta*, 129(3):651–654.

Samel, S. A., Fernandez-Cid, A., Sun, J., Riera, A., Tognetti, S., Herrera, M. C., Li, H., and Speck, C. (2014). A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA. *Genes Dev.*, 28(15):1653–1666.

Sanchez-Pulido, L., Diffley, J. F., and Ponting, C. P. (2010). Homology explains the functional similarities of Treslin/Ticrr and Sld3. *Curr. Biol.*, 20(12):R509–510.

Sanchez-Pulido, L. and Ponting, C. P. (2011). Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics*, 27(14):1885–1888.

Sansam, C. G., Goins, D., Siefert, J. C., Clowdus, E. A., and Sansam, C. L. (2015). Cyclin-dependent kinase regulates the length of S phase through TICRR/TRESLIN phosphorylation. *Genes Dev.*, 29(5):555–566.

Sansam, C. G., Pietrzak, K., Majchrzycka, B., Kerlin, M. A., Chen, J., Rankin, S., and Sansam, C. L. (2018). A mechanism for epigenetic control of DNA replication. *Genes Dev.*, 32(3-4):224–229.

Sansam, C. L., Cruz, N. M., Danielian, P. S., Amsterdam, A., Lau, M. L., Hopkins, N., and Lees, J. A. (2010). A vertebrate gene, ticrr, is an essential checkpoint and replication regulator. *Genes Dev.*, 24(2):183–194.

Sato, M., Gotow, T., You, Z., Komamura-Kohno, Y., Uchiyama, Y., Yabuta, N., Nojima, H., and Ishimi, Y. (2000). Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex. *J. Mol. Biol.*, 300(3):421–431.

Scheres, S. H. (2016). Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Meth. Enzymol.*, 579:125–157.

Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., and Jasin, M. (2011). Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell*, 145(4):529–542.

Schlosshauer, M. and Baker, D. (2004). Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Sci.*, 13(6):1660–1669.

Schneider, J., Bajwa, P., Johnson, F. C., Bhaumik, S. R., and Shilatifard, A. (2006). Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.*, 281(49):37270–37274.

Sclafani, R. A. and Holzen, T. M. (2007). Cell cycle regulation of DNA replication. *Annu. Rev. Genet.*, 41:237–280.

Semlow, D. R., Zhang, J., Budzowska, M., Drohat, A. C., and Walter, J. C. (2016). Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell*, 167(2):498–511.

Sengupta, S., van Deursen, F., de Piccoli, G., and Labib, K. (2013). Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr. Biol.*, 23(7):543–552.

Sheu, Y. J. and Stillman, B. (2006). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol. Cell*, 24(1):101–113.

Sheu, Y. J. and Stillman, B. (2010). The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature*, 463(7277):113–117.

Siddiqui, K., On, K. F., and Diffley, J. F. (2013). Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol*, 5(9).

Simon, A. C., Sannino, V., Costanzo, V., and Pellegrini, L. (2016). Structure of human Cdc45 and implications for CMG helicase function. *Nat Commun*, 7:11638.

Simon, A. C., Zhou, J. C., Perera, R. L., van Deursen, F., Evrin, C., Ivanova, M. E., Kilkenny, M. L., Renault, L., Kjaer, S., Matak-Vinkovic, D., D., Labib, K., Costa, A., and Pellegrini, L. (2014). A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature*, 510(7504):293–297.

Siow, C. C., Nieduszynska, S. R., Muller, C. A., and Nieduszynski, C. A. (2012). OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.*, 40:D682–686.

Smith, D. J. and Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature*, 483(7390):434–438.

Smith, K. D., Fu, M. A., and Brown, E. J. (2009). Tim-Tipin dysfunction creates an indispensible reliance on the ATR-Chk1 pathway for continued DNA synthesis. *J. Cell Biol.*, 187(1):15–23.

Some, D., Amartely, H., Tsadok, A., and Lebendiker, M. (2019). Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *J Vis Exp*, (148).

Sparks, J. L., Chistol, G., Gao, A. O., Raschle, M., Larsen, N. B., Mann, M., Duxin, J. P., and Walter, J. C. (2019). The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell*, 176(1-2):167–181.

Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979). Isolation and characterisation of a yeast chromosomal replicator. *Nature*, 282(5734):39–43.

Stingele, J., Bellelli, R., Alte, F., Hewitt, G., Sarek, G., Maslen, S. L., Tsutakawa, S. E., Borg, A., Kjaer, S., Tainer, J. A., Skehel, J. M., Groll, M., and Boulton, S. J. (2016). Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol. Cell*, 64(4):688–703.

Stodola, J. L. and Burgers, P. M. (2016). Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nat. Struct. Mol. Biol.*, 23(5):402–408.

Sultana, A. and Lee, J. E. (2015). Measuring protein-protein and protein-nucleic Acid interactions by biolayer interferometry. *Curr Protoc Protein Sci*, 79:1–26.

Sun, J., Fernandez-Cid, A., Riera, A., Tognetti, S., Yuan, Z., Stillman, B., Speck, C., and Li, H. (2014). Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function. *Genes Dev.*, 28(20):2291–2303.

Sun, J., Shi, Y., Georgescu, R. E., Yuan, Z., Chait, B. T., Li, H., and O'Donnell, M. E. (2015). The architecture of a eukaryotic replisome. *Nat. Struct. Mol. Biol.*, 22(12):976–982.

Szambowska, A., Tessmer, I., Kursula, P., Usskilat, C., Prus, P., Pospiech, H., and Grosse, F. (2014). DNA binding properties of human Cdc45 suggest a function as molecular wedge for DNA unwinding. *Nucleic Acids Res.*, 42(4):2308–2319.

Szyjka, S. J., Viggiani, C. J., and Aparicio, O. M. (2005). Mrc1 is required for normal progression of replication forks throughout chromatin in S. cerevisiae. *Mol. Cell*, 19(5):691–697.

Tahirov, T. H. (2012). Structure and function of eukaryotic DNA polymerase $\delta$. *Subcell. Biochem.*, 62:217–236.

Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev.*, 17(9):1153–1165.

Tanaka, H., Katou, Y., Yagura, M., Saitoh, K., Itoh, T., Araki, H., Bando, M., and Shirahige, K. (2009). Ctf4 coordinates the progression of helicase and DNA polymerase alpha. *Genes Cells*, 14(7):807–820.

Tanaka, S. and Araki, H. (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb Perspect Biol*, 5(12):a010371.

Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., and Araki, H. (2011a). Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr. Biol.*, 21(24):2055–2063.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature*, 445(7125):328–332.

Tanaka, T., Umemori, T., Endo, S., Muramatsu, S., Kanemaki, M., Kamimura, Y., Obuse, C., and Araki, H. (2011b). Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. *EMBO J.*, 30(10):2019–2030.

Tarsounas, M., Davies, D., and West, S. C. (2003). BRCA2-dependent and independent formation of RAD51 nuclear foci. *Oncogene*, 22(8):1115–1123.

Taylor, M. R. G. and Yeeles, J. T. P. (2018). The Initial Response of a Eukaryotic Replisome to DNA Damage. *Mol. Cell*, 70(6):1067–1080.

Tercero, J. A., Labib, K., and Diffley, J. F. (2000). DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *EMBO J.*, 19(9):2082–2093.

Thorslund, T. and West, S. C. (2007). BRCA2: a universal recombinase regulator. *Oncogene*, 26(56):7720–7730.

Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J., and Bell, S. P. (2015). Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell*, 161(3):513–525.

Tourriere, H., Versini, G., Cordon-Preciado, V., Alabert, C., and Pasero, P. (2005). Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. *Mol. Cell*, 19(5):699–706.

Truong, L. N. and Wu, X. (2011). Prevention of DNA re-replication in eukaryotic cells. *J Mol Cell Biol*, 3(1):13–22.

Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., Guenther, G., Tomlinson, C. G., Thompson, P., Sarker, A. H., Shen, B., Cooper, P. K., Grasby, J. A., and Tainer, J. A. (2011). Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*, 145(2):198–211.

Turchi, J. J., Huang, L., Murante, R. S., Kim, Y., and Bambara, R. A. (1994). Enzymatic completion of mammalian lagging-strand DNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, 91(21):9803–9807.

Valton, A. L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J. F., and Prioleau, M. N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.*, 33(7):732–746.

Van der Merwe, P. (2001). *Surface Plasmon Resonance in Protein-Ligand interactions: hydrodynamics and calorimetry*. Oxford University Press: Oxford, UK.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T. J., and Walter, J. C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.*, 17(15):1894–1908.

Vijayraghavan, S. and Schwacha, A. (2012). The eukaryotic Mcm2-7 replicative helicase. *Subcell. Biochem.*, 62:113–134.

Villa, F., Simon, A. C., Ortiz Bazan, M. A., Kilkenny, M. L., Wirthensohn, D., Wightman, M., Matak-Vinkovic, D., Pellegrini, L., and Labib, K. (2016). Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. *Mol. Cell*, 63(3):385–396.

Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B. J., and Grunstein, M. (2002). Histone acetylation regulates the time of replication origin firing. *Mol. Cell*, 10(5):1223–1233.

Warren, E. M., Vaithiyalingam, S., Haworth, J., Greer, B., Bielinsky, A. K., Chazin, W. J., and Eichman, B. F. (2008). Structural basis for DNA binding by replication initiator Mcm10. *Structure*, 16(12):1892–1901.

Wasserman, M. R., Schauer, G. D., O'Donnell, M. E., and Liu, S. (2018). Replisome preservation by a single-stranded DNA gate in the CMG helicase. *bioRxiv*.

Wellinger, R. J. and Zakian, V. A. (2012). Everything you ever wanted to know about Saccharomyces cerevisiae telomeres: beginning to end. *Genetics*, 191(4):1073–1105.

Whitehouse, I. and Smith, D. J. (2013). Chromatin dynamics at the replication fork: there's more to life than histones. *Curr. Opin. Genet. Dev.*, 23(2):140–146.

Wilmes, G. M., Archambault, V., Austin, R. J., Jacobson, M. D., Bell, S. P., and Cross, F. R. (2004). Interaction of the S-phase cyclin Clb5 with an "RXL" docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes Dev.*, 18(9):981–991.

Wohlschlegel, J. A., Dwyer, B. T., Dhar, S. K., Cvetic, C., Walter, J. C., and Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science*, 290(5500):2309–2312.

Wu, M., Lu, W., Santos, R. E., Frattini, M. G., and Kelly, T. J. (2014). Geminin inhibits a late step in the formation of human pre-replicative complexes. *J. Biol. Chem.*, 289(44):30810–30821.

Wu, R. A., Semlow, D. R., Kamimae-Lanning, A. N., Kochenova, O. V., Chistol, G., Hodskinson, M. R., Amunugama, R., Sparks, J. L., Wang, M., Deng, L., Mimoso, C. A., Low, E., Patel, K. J., and Walter, J. C. (2019). TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature*, 567(7747):267–272.

Wyatt, P. J. (1993). Light scattering and the absolute characterization of macromolecules. *Analytica chimica acta*, 272(1):1–40.

Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., and Aparicio, O. M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. *Science*, 294(5550):2357–2360.

Yabuki, N., Terashima, H., and Kitada, K. (2002). Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells*, 7(8):781–789.

Yabuuchi, H., Yamada, Y., Uchida, T., Sunathvanichkul, T., Nakagawa, T., and Masukata, H. (2006). Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins. *EMBO J.*, 25(19):4663–4674.

Yan, H., Gibson, S., and Tye, B. K. (1991). Mcm2 and Mcm3, two proteins important for ARS activity, are related in structure and function. *Genes Dev.*, 5(6):944–957.

Yao, N. Y. and O'Donnell, M. (2012). The RFC clamp loader: structure and function. *Subcell. Biochem.*, 62:259–279.

Yeeles, J. T., Deegan, T. D., Janska, A., Early, A., and Diffley, J. F. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*, 519(7544):431–435.

Yu, C., Gan, H., Han, J., Zhou, Z. X., Jia, S., Chabes, A., Farrugia, G., Ordog, T., and Zhang, Z. (2014). Strand-specific analysis shows protein binding at replication forks and PCNA unloading from lagging strands when forks stall. *Mol. Cell*, 56(4):551–563.

Yuan, Z., Bai, L., Sun, J., Georgescu, R., Liu, J., O'Donnell, M. E., and Li, H. (2016). Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nat. Struct. Mol. Biol.*, 23(3):217–224.

Yuan, Z., Riera, A., Bai, L., Sun, J., Nandi, S., Spanos, C., Chen, Z. A., Barbon, M., Rappsilber, J., Stillman, B., Speck, C., and Li, H. (2017). Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat. Struct. Mol. Biol.*, 24(3):316–324.

Zegerman, P. (2015). Evolutionary conservation of the CDK targets in eukaryotic DNA replication initiation. *Chromosoma*, 124(3):309–321.

Zegerman, P. and Diffley, J. F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature*, 445(7125):281–285.

Zegerman, P. and Diffley, J. F. (2010). Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. *Nature*, 467(7314):474–478.

Zellweger, R., Dalcher, D., Mutreja, K., Berti, M., Schmid, J. A., Herrador, R., Vindigni, A., and Lopes, M. (2015). Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *J. Cell Biol.*, 208(5):563–579.

Zhai, Y., Cheng, E., Wu, H., Li, N., Yung, P. Y., Gao, N., and Tye, B. K. (2017a). Open-ringed structure of the Cdt1-Mcm2-7 complex as a precursor of the MCM double hexamer. *Nat. Struct. Mol. Biol.*, 24(3):300–308.

Zhai, Y., Li, N., Jiang, H., Huang, X., Gao, N., and Tye, B. K. (2017b). Unique Roles of the Non-identical MCM Subunits in DNA Replication Licensing. *Mol. Cell*, 67(2):168–179.

Zhang, D. and O'Donnell, M. (2016). The Eukaryotic Replication Machine. *Enzymes*, 39:191–229.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.*, 193(1):1–12.

Zheng, L. and Shen, B. (2011). Okazaki fragment maturation: nucleases take centre stage. *J Mol Cell Biol*, 3(1):23–30.

Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, 14(4):331–332.

Zhou, B. B. and Elledge, S. J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature*, 408(6811):433–439.

Zhu, W., Ukomadu, C., Jha, S., Senga, T., Dhar, S. K., Wohlschlegel, J. A., Nutt, L. K., Kornbluth, S., and Dutta, A. (2007). Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. *Genes Dev.*, 21(18):2288–2299.

Zou, L. and Elledge, S. J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science*, 300(5625):1542–1548.

Zou, L., Mitchell, J., and Stillman, B. (1997). CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication. *Mol. Cell. Biol.*, 17(2):553–563.

# Appendix A

# List of oligonucleotides

## A.1 DNA-binding experiment

The sequences of DNA used for the DNA-binding experiment in chapter 4. Ordered from Sigma Aldrich.

*VAJ-01-EMSA*
5′-[6FAM]CTT GCC CAT CCG TAG AAC C-3′

*VAJ-02-EMSA*
5′-GGT TCT ACG GAT GGG CAA G-3′

VAJ01 and VAJ02 are complementary to each other

GQ and CGQ ordered from IDT.
*GQ*
5′-[6FAM]TGA GGG TGG GTA GGG TGG GTA A-3′

*CGQ*
5′-[6FAM]TGA GAG TGA GTA GAG TGA GTA A C-3′

## A.2 Primers for molecular cloning

### A.2.1 Cdc45 cloning

Primers used for Cdc45 cloning. Ordered from Sigma Aldrich.

*hCdc45-Fw-NdeI*
5′-GGA ATT CCA TAT GTT CGT GTC CGA TTT C-3′
*hCdc45-Rev-XhoI*
5′-GGA TCC CTC GAG TTA GGA CAG GAG GGA AAT AAG T-3′
*hCdc45-k420a-qcfw*
5′-GCC TGG AAC TCG CCG CGA AGC AGC TGC GAG-3′
*hCdc45-k420a-qcrev*
5′-CTC GCA GCT GCT TCG CGG CGA GTT CCA GGC-3′
*hCdc45-k421a-qcfw*
5′-GGA ACT CGC CAA GGC GCA GCT GCG AGC C-3′
*hCdc45-k421a-qcrev*
5′-GGC TCG CAG CTG CGC CTT GGC GAG TTC C-3′
*hCdc45-r424a-qcfw*
5′-CAA GAA GCA GCT GGC AGC CAC CCA GCA G-3′
*hCdc45-r424a-qcrev*
5′-CTG CTG GGT GGC TGC CAG CTG CTT CTT G-3′

### A.2.2 Treslin cloning

Primers used for Treslin cloning. Ordered from Sigma Aldrich.
*hTreslin626-NcoI-fw*
5′-AGG GTT CCA TGG GCA CCC TGC GTA GCA G-3′
*hTreslin813-XhoI-rev*
5′-GAA CCG CTC GAG TTA GCT ATC ATC ACT GAA AAA ATC G-3′
*hTreslin503-NcoI-fw*
5′-TTC ATG CCA TGG CCT GTT GCC ATA A-3′
*hTreslin934-XhoI-rev*
5′-GAA CCG CTC GAG TTA AGG CAG ACC ACG TTT C-3′
*hTreslin626-EcoRI-fw*
5′-ACA CGA ATT CAC CCT GCG TAG CAG C-3′
*hTreslin813-XhoI-rev2*
5′-CGC ACT CGA GTTA GTG GTG ATG GTG ATG ATG GCT ATC ATC ACT GAA AAA ATC-3′
*hTreslin881-XhoI-rev*
5′-CGC ACT CGA GTT AAT GGT GGT GGT GAT GGT GAT GAT GAA CTT TCG GAA TTT CAA TCT GG-3′

*hTreslin989-XhoI-rev*
5′-CGC ACT CGA GTT AAT GGT GGT GGT GAT GGT GAT GGT GTG AGC TAC GAC CTT TAA TCT-3′

*hTreslin1049-XhoI-rev*
5-′CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GAT CCT GCT GGA AGC TAT G-3′

*hTreslin1053-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GCT GAT CGC TTT TAT CCT G-3′

*hTreslin1056-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GAT TTT CAC GCT GAT CGC-3′

*hTreslin1059-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GAA CCG GTG AAT TTT CA-3′

*hTreslin1062-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GAA TGC TCT GAA CCG GT-3′

*hTreslin1065-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GAG GGC TAC GAA TGC TCT-3′

*hTreslin1068-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GCA GAC TTT TAG GGC TAC G-3′

*hTreslin1071-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GAC CAA ACA GCA GAC TTT TAG G-3′

*hTreslin1074-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GGC TCA TTG CAC CAA ACA-3′

*hTreslin1077-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GAA TCA TTT CGC TCA TTG C-3′

*hTreslin1080-XhoI-rev*
5′-CGC ACT CGA GTT AGT GGT GAT GGT GAT GAT GGC TAG GGC TAA TCA TTT CG-3′

*hTreslin813-EcoRI-fw*
5′-ACA CGA ATT CGG CAG CAG CAT GAC CCA AGA GAA TAA-3′

*Avi-NcoI-fw*

5′-GGC ACC ATG GAT ATC GAA TTC GGA CTG AAT GAC ATT TTC GAA GCA CAG AAG ATC GAA TGG CAT GAA GGG AGT ATG GAT AAA ATC GAA GAA GGT A-3′

*hTreslin-s820a-qcfw*

5′-GAC CCA AGA GAA TAA AGC GCC GCT GCT GAG CGT TCC-3′

*hTreslin-s820a-qcrev*

5′-GGA ACG CTC AGC AGC GGC GCT TTA TTC TCT TGG GTC-3′

*hTreslin-s838a-qcfw*

5′-GTC GTA GCG TTA GCG GCG CTC CGG AAA GTG ATG AAC-3′

*hTreslin-s838a-qcrev*

5′-GTT CAT CAC TTT CCG GAG CGC CGC TAA CGC TAC GAC-3′

*hTreslin-s923a-qcfw*

5′-TAATCAAGAGCTGCTGGCGCCGAGCAAACGTAGTCTG-3′

*hTreslin-s923a-qcrev*

5′-CAG ACT ACG TTT GCT CGG CGC CAG CAG CTC TTG ATT A-3′

*hTreslin-s969a-qcfw*

5′-CAA AAG CGT TGC CGA AGC ACC GGT GCA TAA ACA AAT C-3′

*hTreslin-s969a-qcrev*

5′-GAT TTG TTT ATG CAC CGG TGC TTC GGC AAC GCT TTT G-3′

*hTreslin-s1001a-qcfw*

5′-GGT GTT GTT GAA GAA GCT CCG GAA AAA GGT GAC G-3′

*hTreslin-s1001a-qcrev*

5′-CGT CAC CTT TTT CCG GAG CTT CTT CAA CAA CAC C-3′

*hTreslin-s1013a-qcfw*

5′-TAG TCT GCG TCGT GCA CCG CGT ATT AAA CAG CTG-3′

*hTreslin-s1013a-qcrev*

5′-CAG CTG TTT AAT ACG CGG TGC ACG ACG CAG ACT A-3′

*hTreslin-t969ps-qcfw*

5′-CAA AAG CGT TGC CGA ATA GCC GGT GCA TAA ACA AAT C-3′

*hTreslin-t969ps-qcrev*

5′-GAT TTG TTT ATG CAC CGG CTA TTC GGC AAC GCT TTT G-3′

*hTreslin-s1001ps-qcfw*

5′-GGT GTT GTT GAA GAA TAG CCG GAA AAA GGT GAC G-3′

*hTreslin-s1001ps-qcrev*

5′-CGT CAC CTT TTT CCG GCT ATT CTT CAA CAA CAC C-3′

*hTreslin-t969e-qcfw*
5′-CAA AAG CGT TGC CGA AGA ACC GGT GCA TAA ACA AAT C-3′
*hTreslin-t969e-qcrev*
5′-GAT TTG TTT ATG CAC CGG TTC TTC GGC AAC GCT TTT G-3′
*hTreslin-s1001d-qcfw*
5′-GGT GTT GTT GAA GAA GAT CCG GAA AAA GGT GAC G-3′
*hTreslin-s1001d-qcrev*
5′-CGT CAC CTT TTT CCG GAT CTT CTT CAA CAA CAC C-3′

### A.2.3 TopBP1 cloning

Primers used for TopBP1 cloning. Ordered from Sigma Aldrich.
*hTopBP1-NcoI-fw*
5′-CGA CCA TGG CGA TGT CCA GAA ATG ACA AAG A-3′
*hTopBP1-290-SalI-rev*
5′-ACG CGT CGA CTT ATG CTT CTG GTC TAG GTT C-3′

### A.2.4 BirA cloning

Primers used for BirA cloning. Ordered from Sigma Aldrich.
*MBP-BirA-EcoRI-fw*
5′-CGC GGA ATT CAT GAA AAT CGA AGA AGG TAA ACT-3′
*BirA-SalI-rev*
5′-GTC AGT CGA CTT ATT ATT TTT CTG CAC TAC GCA GG-3′
*Strep-MBP-BirA-EcoRI-fw*
5′-GTC AGA ATT CAT GTG GTC CCA CCC CCA GTT CGA AAA ACA TCA CCA TCA CCA TCA CAA C-3′

### A.2.5 MCM cloning

Primers used for MCM cloning. Ordered from Sigma Aldrich.
*StrepTagII-Fw-NgoMIV*
5′-ATA TGC CGG CGC GCC ACC ATG AG-3′
*StrepTagII-Rev-NgoMIV*
5′-ATA TGC CGG CCC CTT GAA AAT ACA GGT TCT CC-3′
*StrepTagII-Fw-SalI*
5′-TAT AGT CGA CGC CAC CAT GAG CG′-3′

*StrepTagII-Rev-SalI*

5′-TAT AGT CGA CCC TTG AAA ATA CAG GTT CTC C-3′

*StrepII-String*

5′-ATG ACT CCC GGG CTC GAG CGC CAC CAT GAG CGC CTG GTC CCA CCC CCA GTT CGA AAA AGG CGG AGG CTC TGG CGG CGG AAG CGG AGG ATC TGC TTG GAG CCA CCC TCA GTT TGA GAA GGA GAA CCT GTA TTT TCA AGG CTC GAG ATG ACT-3′

## A.3 Vector map

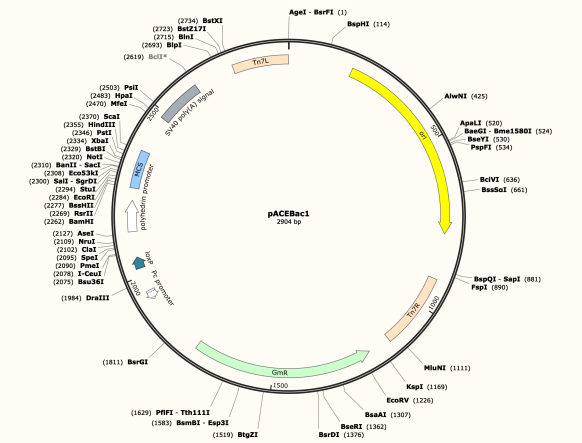



**Fig. A.1 Vector map for pACEBac1**

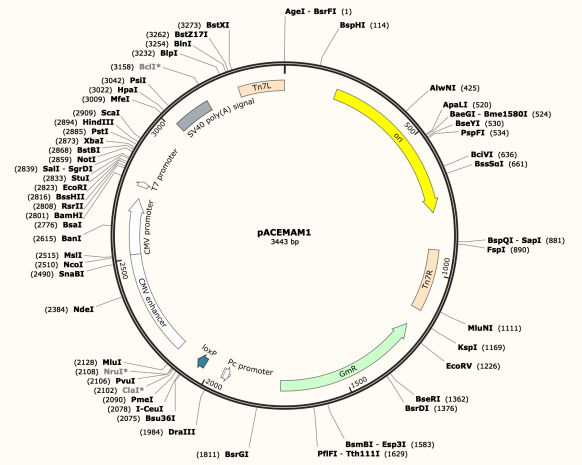



**Fig. A.2 Vector map for pACEMam1**

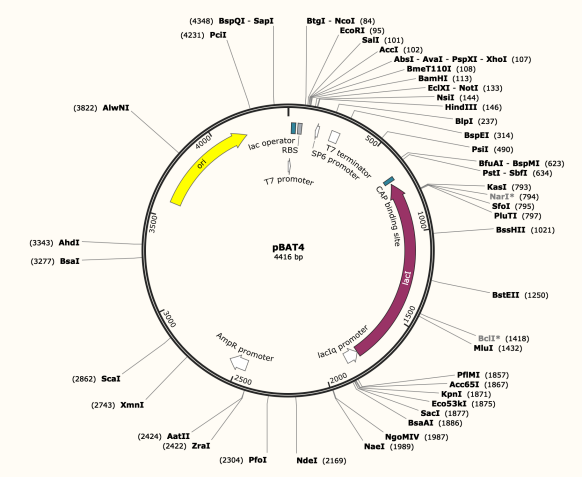



**Fig. A.3 Vector map for pBAT4**

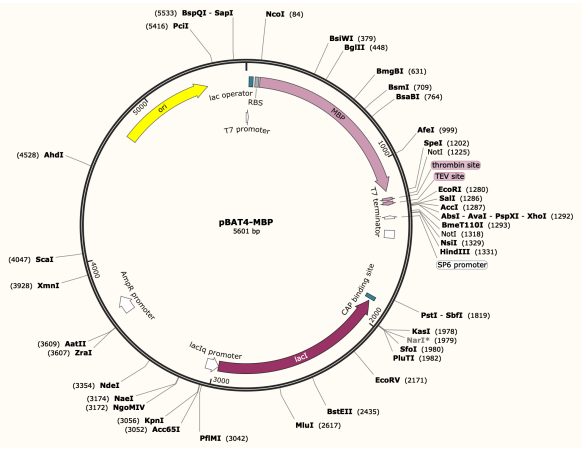



**Fig. A.4 Vector map for pBAT4-MBP**

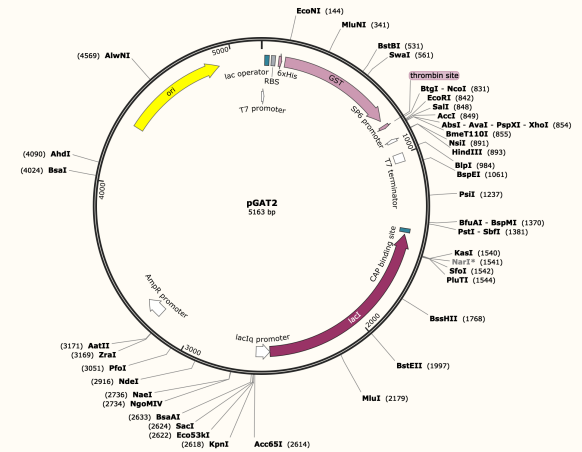



**Fig. A.5 Vector map for pGAT2**

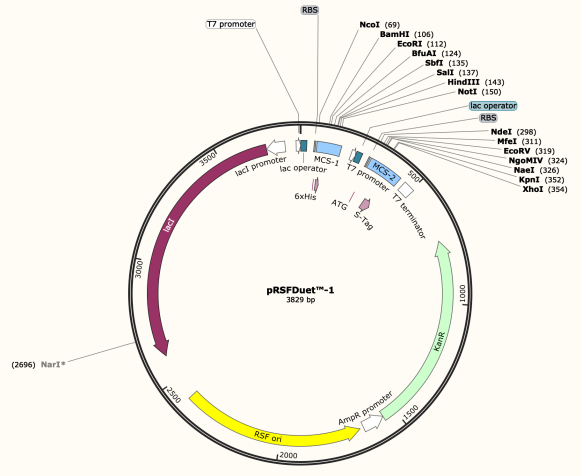



**Fig. A.6 Vector map for pRSFDuet1**

# Appendix B

# SEC-MALS of MBP control

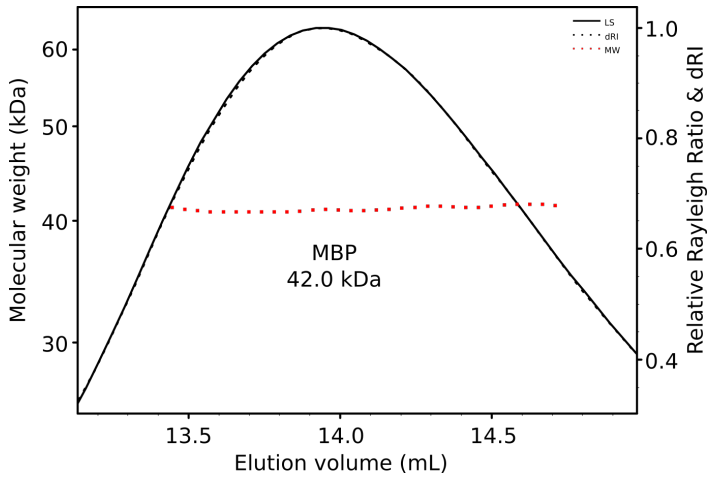



**Fig. B.1 SEC-MALS of MBP as a control.** MBP as a control for all SEC-MALS experiments described in chapter 3. Eluted as a single peak at 42.0 kDa. Polydispersity = 1.000 ($\pm$0.210%). Predicted Mw = 43.0 kDa

SEC-MALS of MBP control was performed as follows. Purified MBP (expressed from empty vector of pBAT4-MBP) at 2 mg/mL was applied to Superdex200 Increase 10/300 column (GE Healthcare) using buffer 20 mM HEPES-NaOH pH 8.0, 160 mM NaCl, 1 mM DTT at flow rate of 0.5 mL/min.

# Appendix C

# Cryo-EM data collection parameters

**Table C.1** Cryo-EM data collection parameters

| **Parameter** | **Cdc45–Treslin$_{SHD}$** | **Cdc45–Treslin$_{SHD}$ crosslinked** | **Cdc45–Treslin$_{SHD\text{-}ext}$** | **apo-MCM** | **MCM + ATP** | **MCM–Treslin$_{SHD\text{-}ext}$** |
|---|---|---|---|---|---|---|
| Nominal magnification (×) | 215,000 | 215,000 | 165,000 | 73,000 | 73,000 | 73,000 |
| Voltage (kV) | 300 | 300 | 300 | 300 | 300 | 300 |
| Detector | K2 counting | K2 counting | K2 counting | K2 counting | Falcon II linear | K2 counting |
| Frames | 40 | 40 | 35 | 80 | 40 | 40 |
| Dose per frame ($e^-/A^2$) | 1.28 | 1.29 | 2.32 | 0.66 | 0.66 | 0.88 |
| Defocus range ($\mu$m) | -0.5, -0.8 | -0.5, -0.8 | -0.75, -1.0 | -1.5 to -3.0 | -1.5 to -3.0 | -1.5 to -3.0 |
| Pixel size (Å/pixel) | 0.66 | 0.66 | 0.85 | 1.047 | 1.047 | 1.048 |
| Symmetry | C1 | C1 | C1 | C1 | C1 | C1 |
| Micrographs collected | 1,521 | 1,592 | 1,170 | 3,350 | 1,350 | 3,171 |
| Initial particles | 393,780 (393,210) | 398,556 (339,783) | 132,610 | 644,719 | 88,371 | 533,326 |
| Final particles | 78,788 (71,293) | 25,556 (50,314) | 24,287 | 65,655 | 25,300 | 44,300 |
| Map resolution (Å) | 10 (8.8) | 12 (11) | 10 | 6.8 | 10.1 | 8.2 |
| Map resolution range (Å) | 9–12 (9–12) | 11–14 (10–14) | 9–12 | 6–9 | 8–12 | 8–14 |
| Volta phase plate | Yes | Yes | Yes | No | No | No |
| Location of data collection | Dept of Biochemistry, Cambridge | Dept of Biochemistry, Cambridge | eBIC, Didcot | eBIC, Didcot | Dept of Biochemistry, Cambridge | eBIC, Didcot |

Notes: the parameters in parentheses were obtained during data processing in cryoSPARC 2.8.0 software packages